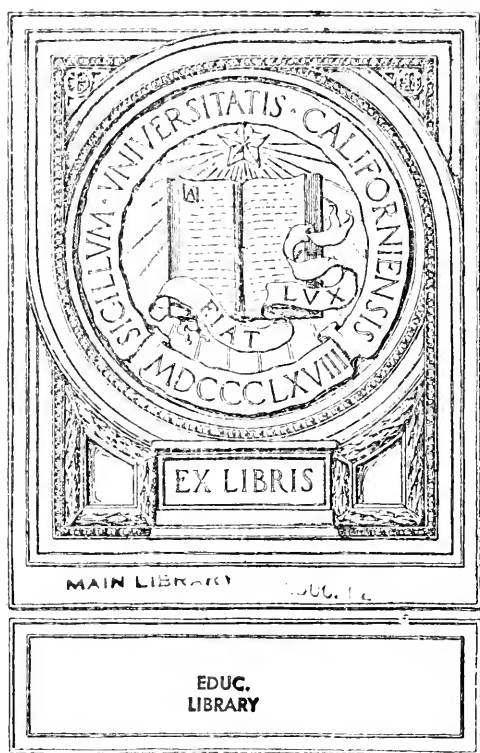# Columbia University

Contributions to Education

Teachers College Series

A Study in Educational Prognosis

By

Elbert Kirtley Fretwell, Ph.D.

ACKNOWLEDGMENTS

CONTENTS

A Study in Educational Prognosis

CHAPTER I

THE PROBLEM

CAN academic success be predicted? If so, how? Three bases of prognosis have received much consideration: College entrance examinations, teachers' estimates, and school marks.

In 1906, Thorndike [1] pointed out that there was a low correlation between the marks of pupils in college entrance examination and their marks later in college. Adam Leroy Jones, as champion of the college entrance examination plan, maintained that "No advocate of examinations ever supposed that the purpose of examinations was to furnish a prediction of what the boy would do . . . through his college course, or indeed even through the first year of the course." [2] It is certain that teachers' estimates are not perfect in selecting the pupils who can pass the examinations of the College Entrance Examination Board. For example, in 1916, three fourths of the students specially recommended by their teachers as able to pass the examinations in American history, in mediæval and modern history, and in civil government, failed to make a grade of sixty per cent. [3] Can twenty-five per cent efficiency in estimating academic success be considered successful?

Do school marks foretell academic success better than do teachers' estimates? In spite of the different standards of marking of different schools, of different departments, and by different teachers, of the different emphasis placed upon different parts of the same work, of the inability of some teachers to see small differences—in spite of all these differences, are school marks a

[1] "Future of the College Entrance Examination Board," *Educational Review*, 31:5.
[2] "Entrance Examination and College Records," *Educational Review*, 48:109, 1914.
[3] *Sixteenth Annual Report of College Entrance Examination Board*, 1916.

more accurate basis for prognosis than teachers' estimates? Without discussing whether Dearborn's coefficient of average or average of coefficients is the more suitable for his work, the conclusion that he reaches may be noted: In seventy-five per cent of the cases the standing in the university can be predicted from the standing in the high school.[4]   F. O. Smith found that with 120 students at the University of Iowa there was a correlation of .53 between the average of all high school marks and all marks in the university.[5]   Walter W. Pettit found a correlation of .63 between the average of all high school marks and the freshman marks in college.[6]   In the cases of 253 Harvard students, E. A. Lincoln found that the correlation between high school standing and standing in the college entrance examination was .46, the correlation between college entrance examination and standing the freshman year in college, .47, while the correlation between high school standing and freshman college standing was .69. Therefore, he concludes that school marks furnish a better basis for prognosis than entrance examinations.[7]

Can school marks be considered accurate when the marks of 142 English teachers, as Starch and Elliot have pointed out, vary in grading the same composition from 50 to 98,[8] and the marks of 118 mathematics teachers for the same paper in mathematics vary from 28 to 90?[9]   Some recognition of this wide differing is necessary in order to appreciate the extraordinary variability in teachers' marks pointed out by F. J. Kelly.[10]

Unreliable as school marks may be, Truman Lee Kelley found that, for estimating the pupil's scholastic ability, the elementary school records of the pupil gave more accurate information than either the teachers' estimates or the tests he devised.[11]

[4] Bulletin No. 312, High School Series No. 6, University of Wisconsin, 1909.

[5] *A Rational Basis for Determining Fitness for College Entrance*, University of Iowa Studies, Vol. 1, No. 3, 1910.

[6] *A Comparative Study of New York High School and Columbia College Grades*, Master's Essay, Teachers College, 1912.

[7] *School and Society*, Vol. V, No. 119, p. 417, 1917.

[8] *School Review*, 20:442-457.

[9] *Ibid.*, 21:254-259.

[10] *Teachers' Marks*, Teachers College, Contributions to Education, 1914.

[11] *Educational Guidance*, Teachers College, Contributions to Education, 1914.

When examinations, teachers' estimates, and school marks have been considered, is there any other basis for predicting academic success? Standardized educational and psychological tests form the basis in this study for predicting a pupil's success. On the basis of standardized tests, to what extent can a pupil's academic success be foretold? An answer to this question constitutes the theme of the work that follows—A Study in Educational Prognosis.

CHAPTER II

THE EXPERIMENT

1. Conditions Under Which the Experiment Was Made

This study in educational prognosis concerns itself chiefly with an experiment in organizing into homogeneous groups the pupils who entered a junior high school. The experiment has been made possible by Teachers College, Columbia University, and the public school system of New York City coöperating in the organization of the Speyer School as an experimental academic junior high school for boys. This school, as a part of the free public school system of New York City, opened February, 1916, with about two hundred boys who had finished the first six grades of the regular schools. One hundred additional boys entered in September, 1916, and fifty more entered in February, 1917. It is the first group—the one entering in February, 1916 —that forms the basis of this study.

When the two hundred pupils entered the school, an attempt was made to organize them for purposes of instruction into homogeneous groups. On account of the size of the class rooms, the groups were limited to twenty-five pupils each. All groups, when so organized, were to follow the same course of study, but each group was to proceed as rapidly as it was able, i.e., at its optimum speed. This means that the abler classes, with the revised and enriched course of study, with the improved method of instruction and of study, have the opportunity of completing the three years' work of the junior high school as rapidly as they are able—possibly in two years. If such is the case, the pupils so doing will pass from the completion of the 6B grade to the second year of the senior high school in two years. The virtue of the plan lies partly in the fact that the brighter pupils are not held back by the slower ones, and that these slower pupils are not discouraged by being rushed beyond their best rate of

work or by being placed in groups with pupils with whom they cannot compete.

One of the first problems in organizing the school with this limited number of pupils was to find to what extent the pupils were typical 6B boys. Before the opening of the school in February, 1916, it was found that pupils would be sent from five public schools—Numbers 5, 10B, 43, 184, and 186 Manhattan. Accordingly, each teacher of each 6B class in these five schools was asked to rank his or her pupils separately in intelligence and in industry. In addition there was given to all of these 6B boys, the Woody Multiplication Scale, the Trabue Completion-Test Language Scales B and C, and fifty words from the Ayres Spelling Scale, list Q. Each pupil also wrote an English composition on the subject, ''How I Should Spend Twenty Dollars.'' Thus two standards were provided by means of which it was possible to compare those pupils who came to Speyer School with all the other boys of the twenty-four classes from which they came. This was evidently necessary in order that one might know to what extent the pupils included in this study were typical 6B boys.

While the working out of this preliminary problem was necessary, the real problem was to classify into homogeneous groups, on the basis of mental ability, all the pupils present at the opening of the school. To do this the scores were retained that these pupils had made in the five tests—Woody Multiplication, Ayres Spelling, Composition, Trabue Completion-Test Language Scales B and C—and six additional tests were then given: Thorndike Reading Alpha 2, Part II, Thorndike Visual Vocabulary A, and Woodworth and Wells Easy Opposites, Easy Directions and Mixed Relations. Each of these tests was chosen because it had shown in previous experiments a positive correlation with desirable traits. When the achievement of all pupils in each test had been ranked and each pupil's ranks in all tests had been added, it was possible by ranking these totals to state in a single figure where, on the basis of achievement in the eleven tests, each pupil stood in relation to each of the others of the whole group. The pupil with the highest score on the basis of the tests was ranked one, the second best, two, and so on. Those

ranking from one to twenty-five were placed in the first class, A1; pupils twenty-six to fifty were placed in A2; pupils fifty-one to seventy-five in A3, and so on to pupils one hundred seventy-six to two hundred in A8. A pupil was not, however, fixed finally according to his original grouping. Whenever the teachers of any pupil agreed that he was in too slow or too fast a group, he was transferred. The fact that there were several groups made it possible for the teachers to make these transfers and still keep the classes about the original size.

The criterion of prognosis in this experiment must rest finally in the teachers' judgments. By keeping a record of all transfers from one group to another, and by having the teachers rank the pupils after teaching them one year, it is possible to see how nearly the classification by the tests in the beginning corresponds to that made by the teachers after teaching the pupils one year.

In addition to the amount of statistical work involved, there were some limitations on including in this study all of the two hundred pupils of the first group that entered Speyer School. Due to absence from school when the first five tests were given in the twenty-four class rooms of the five public schools, some pupils missed one or more of the tests. There were in all ninety-seven pupils who had scores in every one of the eleven tests. Of these ninety-seven, seventy-four were still in school at the end of one year. Fortunately for this study, these seventy-four pupils were scattered through all of the groups from the fastest to the slowest. As a result of the departmental plan of teaching there were, aside from the teachers of drawing, music, shop, gymnasium, and general science, four teachers of regular academic subjects who were teaching all of these seventy-four boys. These teachers, at the end of one year, ranked these pupils for general mental ability. When all cases have been considered, it will be seen that marks made in class correspond very closely to the estimate given by the teachers, yet each teacher was asked to rank the pupils on his or her own definition of general mental ability, and to make the ranking without consulting anyone. This was done.

The data for this study therefore consist of the scores of all

the 6B boys in twenty-four classes in five New York City public schools, in five standardized tests; the ranking by each teacher of these twenty-four classes, of the boys of his or her class for intelligence and for industry; the scores of about one hundred and seventy-five pupils drawn from these twenty-four classes of 6B boys, in eleven educational and psychological tests; the record of all transfers from the grouping according to these tests made by the Speyer teachers during the first year of their teaching these pupils; all school marks of seventy-four pupils of the first six grades of the public schools; all school marks of the same group during their first year in the junior high school; the age of the seventy-four pupils; the ranking of seventy-four pupils for whom there were scores in eleven tests and who remained in school one year, by four teachers at the end of that year. In addition, the eleven tests were repeated at the end of one year, i.e., the same or similar tests were given to the seventy-four boys.

2. The Tests

Achievement in standardized educational and psychological tests was the basis for organizing the pupils into groups for purposes of instruction. The size of the class rooms limited these groups to twenty-five pupils each.

Eleven tests were given in February, 1916, and a like number of the same or of similar tests were given to the same pupils one year later. The tests in 1916 were:

1. Thorndike Reading Scale A, Visual Vocabulary.[1]
2. Thorndike Scale Alpha 2, For Measuring the Understanding of Sentences, Part II.[2]
3. An English composition on the subject, "How I Would Spend Twenty Dollars." [3]
4. Fifty words from the "Q" list of the Ayres Measuring Scale for Ability in Spelling.[3]
5-6. Trabue Completion-Test Language Scales B and C.[4]

[1] Thorndike, "Reading Scale A, Visual Vocabulary," in *Teachers College Record*, September, 1914.
[2] "Scale Alpha 2. For Measuring the Understanding of Sentences," in *Teachers College Record*, Vol. XVI, No. 5, November, 1915.
[3] Ayres, L. P., *Measuring Scale for Ability in Spelling*—Russell Sage Foundation, Division of Education.
[4] Trabue, M. R., *Completion-Test Language Scales*, Teachers College Contributions to Education, No. 77.

7-8. Woody Arithmetic Scales, Multiplication and Division, Series A.[5]

9-10. Woodworth-Wells Logical Relations Tests—Opposites II (north, south, out), Mixed Relations II (good, bad, long).[6]

11. Woodworth-Wells Easy Directions (Cross out the smallest dot).[6]

In 1917 the tests used were:

1. Thorndike Reading Scale A2: Visual Vocabulary plus steps 11, 11½, 12, 12½ of Scale A2: Provisional Extension.[7]

2. Thorndike Reading Scale Alpha 2, Part II, repeated.[2]

3. An English composition on the subject, "How I Should Like to Spend Next Saturday." [8]

4. Ayres Spelling Scale—fifty words selected from the R, S, T, U, V and W lists.[3]

5-6. Trabue Completion-Test Language Scales J and K.[4]

7-8. Woody Arithmetic Scales, Multiplication and Division, Series B.[5]

9-10. Woodworth-Wells Logical Relations: Opposites I (long, soft, white), and Mixed Relations I (eye, see, ear).[6]

11. Woodworth-Wells Easy Directions (Cross out g in tiger).[6]

It will be noted that eight of these 1917 tests are not repetitions, but are similar to those given in 1916. Likewise it will be noted that Reading Alpha 2, Part II, is a repetition of the same test, and that the Woody tests, Series B, are also a repetition of the 1916 tests, but consist of only about half as many problems. While the footnotes accompanying the enumeration of the tests indicate where the reader who is not already acquainted with the tests may refer to them, some description of them may be of value.

The Visual Vocabulary Test, Reading Scale A, given in 1916, consists of forty-three words, beginning with five easy words of equal difficulty and progressing by steps of five-word groups of increasing difficulty to the last three words which are the most difficult of all. This test, built on the "checking by class" principle, requires that the pupil write the letter F under every word meaning a flower, T under every word meaning something about time, and so on through the eight kinds of words composing the test. The Visual Vocabulary test given in 1917, "Visual Vo-

5 Woody, C., *Measurements of Some Achievements in Arithmetic*, Teachers College Contributions to Education, No. 80.

6 Woodworth-Wells, *Association Tests*, in *Psychological Monographs*, Vol. XIII, No. 5, December, 1911.

7 Thorndike, "Reading Scale A2, Visual Vocabulary" in *Teachers College Record*, November, 1916.

8 Hillegas, M. B., *A Scale for the Measurement of Quality in English Composition by Young People*, Teachers College.

cabulary Scale A2,'' plus the four additional steps taken from ''A2 Provisional Extension,'' consists of one hundred and seventy words. It is an extension and improvement of Reading Scale A, but maintains the same obvious purpose, i.e., to measure how hard words a pupil can read in the sense of understanding their meaning well enough to classify them under the proper headings; as, an animal, a flower, something about time, etc. The time allowed, twenty-five minutes, enabled each pupil to attempt to place the correct letter under each word. In scoring, a credit of one was given for each word lettered correctly. The number of words lettered correctly constituted the score.

Scale Alpha 2, For Measuring the Understanding of Sentences, Part II, consists of eight paragraphs of increasing difficulty. Each paragraph is followed by questions—usually three or four—and the pupils' ability to understand the paragraph is determined by his answers to these questions. In the time allowed, twenty-five minutes, all pupils except the very slowest were able to attempt to answer each question. In scoring, two was given for a correct answer and one for a semi-correct answer.

The grade of the English composition on the subject, ''How I Should Spend Twenty Dollars,'' was determined by averaging the marks given by four to six experienced judges, who in forming their judgments used the Hillegas Scale. The composition, ''What I Should Like to Do Next Saturday,'' was graded in the same manner except that there were four judges instead of four to six. The time allowed the pupil for writing this composition was thirty minutes.

In giving the Ayres Spelling test the regular teacher pronounced the words but did not grade the results. Two credits were given for each word spelled correctly.

The Trabue Completion-Test Language Scales B and C consist of ten mutilated sentences. In each sentence, from the first one which is very easy, through the gradually increasing difficulty of each succeeding step to the last one, which is usually beyond the ability of the pupil, the omitted word or words are to be supplied. While C is somewhat more difficult than B, either test in the opinion of the author, Dr. M. R. Trabue, ''measures a class fairly well, but both taken together give a more accurate

measure of the individual.'' Scales J and K, consisting of seven sentences each, are very much more difficult, and are more equally matched than B and C. However, all four seem to measure the same quality, whatever that quality may be. In scoring, the correct answers published by Dr. Trabue were followed absolutely. Two credits were allowed for each sentence perfectly completed and one credit for each sentence almost perfectly completed. Time: seven minutes for each test.

The Woody Multiplication Scale, Series A, consists of thirty-nine problems. The first problem is as easy as it can be made, but there is a gradual increase in difficulty with each succeeding one. The Multiplication Scale, Series B, consists of twenty problems drawn from Scale A. The Division Scale, Series A, consisting of thirty-six problems, is constructed in the same way as the Multiplication Scale, Series A. Division Scale, Series B, is made up of fifteen problems drawn from Division Scale A. In scoring, one credit was allowed for each correct answer. Time: twenty minutes for Series A and ten minutes for Series B.

The two lists of twenty words each which compose the Opposites Test, make it possible to give two tests, of about equal difficulty, of the same function. The pupils were required to write as rapidly as possible the opposite of the word appearing in the printed list. One credit was given for each correct response. Time: seventy-two seconds.

In the Mixed Relations Test, twenty series of three words each, with a fourth word missing, were given. The pupil was to note the relation of the second word to the first, and then find and write down a word standing in the same relation to the third. The two lists of ''mixed relations'' make possible the repetition of the test without any particular interference from learning or remembering. One credit was given for each word correctly supplied. Time: one hundred and twelve seconds.

The Easy Directions Test makes it possible to find out the pupil's ability and speed in understanding and following certain instructions. The two tests of approximately equal difficulty make the repetition of the test possible. One credit was given for each instruction correctly followed. Time: eighty-two seconds.

3. Were the Subjects Typical 6B Boys?

The first problem that presents itself is really a preliminary one. It is to determine to what extent the seventy-four pupils of this study are typical of pupils finishing the 6B grade in New York City public schools. Fortunately it is possible to know this relationship with a considerable degree of exactness. It will be recalled that all the 6B boys, about seven hundred in number, in twenty-four class rooms of five New York City public schools, were given five tests. In addition, the teachers in each of these twenty-four class rooms ranked his or her boys in intelligence and in industry.

It was then possible by comparing the medians, or, in the case of spelling, the averages, of the achievement of the whole group in the five tests given in these twenty-four rooms of the five public schools with the achievement of the pupils of that group who came to Speyer School, and thus know, on the basis of these tests, to what extent the Speyer pupils concerned in this experiment were typical pupils. By reading Table A under the headings "6B Boys—Five Schools," and "Boys who came to Speyer School," it will be seen that the Speyer group is slightly superior; it achieved about one-third of one point more than the larger group in Trabue B, one-fifth of one point more in Trabue C, about one and one-half points more in Woody Multiplication, about two points more in Composition as represented by the average of the grades given by judgment of from four to six judges who used the Hillegas Scale, and about five and one-fifth

TABLE A
Comparison of Median Achievements

	6B BOYS FIVE SCHOOLS		BOYS WHO CAME TO SPEYER		74 BOYS OF THIS STUDY	
	Cases	Median	Cases	Median	Cases	Median
Trabue B	684	12.78	171	13.17	74	13.41
Trabue C	677	12.58	167	12.77	74	12.05
Woody X	707	31.73	170	33.30	74	33.31
Composition	694	30.39	164	32.36	74	33.9
		Average		Average		Average
Spelling	704	89.02	171	94.21	74	93.9

points more in spelling the fifty words of the Ayres Q list. It is noted then that the Speyer group is, on the basis of achieve-

ment in these five tests, somewhat better than the other group, though only slightly better. It should also be pointed out that the Speyer group did not cluster around the median of achievement and that there were all kinds of pupils, from the brightest to very nearly the dullest. On this point the estimates of the twenty-four teachers are in accord with the tests.

The ranking of his or her boys for intelligence and industry by each teacher of the twenty-four class rooms, does not make an accurate comparison of these pupils possible. Since there is no way of comparing the subjective estimate of one teacher concerning one pupil with a like estimate of another teacher of another pupil, such evaluations have worth in rough groupings only. However, the ranking for industry and for intelligence by each teacher of his or her own pupils when compared with the ranking by achievement in the five tests, was possible. A study of the comparative ranking as made by fourteen of these teachers —selected at random—with that made by the five tests is shown in Table B. Here, for example, teacher number one, who ranked pupils practically the same for intelligence and for industry, has a fairly high correlation, .76 (Pearson formula), between intelligence and industry combined, with the composite of the five tests, while teacher number ten finds little relation between intelligence and industry, .29, and a correlation of only .38 between intelligence and industry combined, with the composite of the tests. A glance at the medians is sufficient to show that the easily checked-up abilities represented by multiplication and spelling have, as a rule, a much closer relation to the teacher's estimate of intelligence and industry than have the abilities measured by English composition and the Trabue Completion-Test Language Scales B and C. However, when the composite of all the tests is considered, the relation between these teachers' ranking for intelligence and industry combined and a composite of these five tests varied from a correlation of .17 to one of .76, with a median of .38 and an average of .48. On account of the ranking of the different groups by different teachers, with no one pupil ranked by any two teachers, it is impossible to present in any one statistical statement the exact relation between the ranking of the Speyer pupils by the teacher and the ranking by the

TABLE B

THE RELATION OF ONE TEACHER'S RANKING TO THAT BY STANDARDIZED TESTS

	Intelligence with Industry	Intelligence with Composite of Tests	Industry with Composite of Tests	Int. and Ind. with Composite of Tests	Spelling with Int. and Ind.	Trabue B and C with Int. and Ind.	Woody X with Int. and Ind.	
Teacher No. 1	1.00	.76	.75	.76	.58	.45	.72	.46
(37 Pupils) P.E....		(.05)	(.05)	(.05)	(.07)	(.09)	(.06)	(.09)
Teacher No. 2	.77	.75	.65	.75	.52	.37	.48	.62
(38 Pupils) P.E....	(.05)	(.05)	(.06)	(.05)	(.08)	(.01)	(.09)	(.07)
Teacher No. 3	.83	.53	.67	.64	.46	.28	.59	.37
(27 Pupils) P.E....	(.06)	(.09)	(.07)	(.08)	(.10)	(.2)	(.08)	(.11)
Teacher No. 4	.81	.61	.46	.58	.21	.18	.29	.60
(42 Pupils) P.E....	(.03)	(.07)	(.08)	(.07)	(.1)	(.1)	(.1)	(.07)
Teacher No. 5	.95	.61	.53	.58	.54	.44	—.03	.44
(48 Pupils) P.E....	(.01)	(.06)	(.07)	(.07)	(.07)	(.08)		(.08)
Teacher No. 6	.90	.49	.54	.54	.31	.32	.27	.53
(40 Pupils) P.E....	(.02)	(.07)	(.07)	(.07)	(.09)	(.09)	(.09)	(.07)
Teacher No. 7	.72	.36	.30	.39	.40	.17	.27	.38
(37 Pupils) P.E....	(.05)	(.10)	(.10)	(.09)	(.09)	(.11)	(.10)	(.09)
Teacher No. 8	.99	.44	.45	.38	.29	.15	.16	.44
(40 Pupils) P.E....		(.08)	(.08)	(.09)	(.10)	(.11)	(.10)	(.08)
Teacher No. 9	.37	.41	.32	.38	.31	.15	.19	.34
(37 Pupils) P.E....	(.09)	(.09)	(.10)	(.09)	(.10)	(.11)	(.11)	(.09)
Teacher No. 10	.29	.53	.21	.38	.40	.06	.05	.61
(43 Pupils) P.E....	(.11)	(.07)	(.10)	(.09)	(.09)	(.11)	(.11)	(.06)
Teacher No. 11	.83	.41	.34	.36	.53	.21	.17	.08
(40 Pupils) P.E....	(.03)	(.08)	(.10)	(.10)	(.08)	(.10)	(.10)	(.11)
Teacher No. 12	.91	.37	.35	.36	.26	—.02	.48	.14
(41 Pupils) P.E....	(.02)	(.09)	(.09)	(.09)	(.10)		(.08)	(.10)
Teacher No. 13	.67	.46	.32	.31	.43	.22	—.17	.21
(32 Pupils) P.E....	(.07)	(.09)	(.10)	(.10)	(.10)	(.11)		(.12)
Teacher No. 14	.72	.13	.20	.17	.13	.03	—.04	.45
(37 Pupils) P.E....	(.05)	(.10)	(.11)	(.11)	(.11)	(.11)		(.09)
Average	.77	.49	.43	.48	.38	.21	.24	.40
Median	.82	.51	.40	.38	.40	.19	.23	.44
Range	1.00—.29	.76—.13	.75—.20	.76—.17	.58—.13	.45——.02	.72——.17	.62—.08

pupils' achievement in five standardized tests. However, it has been shown by the tests that the group coming to Speyer School made a little, but just a very little, better scores in the tests than did the other pupils in the twenty-four class rooms in the five schools from which they came.

The next step is to see how the seventy-four boys who form

the basis for this study compare with the whole group that came to Speyer School. This can be done by comparing their scores in the five tests with the scores of the whole group tested in the twenty-four class rooms, or with the scores of all those who came to Speyer School, or with both. By a study of Table A, it will be seen that the seventy-four boys are slightly inferior to the whole group at Speyer School, and a very little better than the whole group from the five schools. The answer, then, to the first problem is that the achievement of the group of seventy-four boys studied, as shown in the results of the five tests, is a little, but a very little, above the average or the median achievement of all the boys in the twenty-four classes of the five schools from which they came.

4. The Scores

In order to rank the seventy-four boys, for their achievement in the eleven tests, it was necessary to find a single statistical statement that represented this achievement. On the basis of the scores that resulted from the tests of February, 1916 (see Table Y on page 51), each individual could be ranked in each test. The pupil making the highest score was ranked one, and the pupil making the poorest, seventy-four. When each individual had been ranked in each subject, his rankings in the eleven subjects were added, and the result was a column of totals representing the combined rankings of each individual in all tests. The rankings of these totals resulted in a single statistical statement of each individual's achievement in relation to the achievement of the seventy-three others of the group.

The eleven tests were considered at first of equal value. While all of these tests had been used before, so far as the writer knows this combination of them in testing pupils of this age, for this purpose, had not been made. Each of the tests had shown in previous experiments a positive correlation with desirable traits, otherwise it would not have been used; but the relative value of the tests for purposes of practical educational prognosis was largely untested. Any weighting given to any one of these tests in the beginning of this experiment would have been largely guesswork. The guess made here was

that any one test was equal to any other test, and the pupils were ranked on that basis. The value of each test for the purposes of this study is considered later.

It is possible to correlate the ranking that resulted from the eleven tests with ranking by age, by grades made during the first six years of public school attendance, by grades made the first year at Speyer School, and with the ranking by four teachers of academic subjects after teaching the pupils one year.

5. AGE

Age, if taken in years and months and carefully checked, is definite. Following the studies of T. L. Kelley, McCall and others, a negative correlation between age and achievement was to be expected. Hence the youngest pupil was ranked one and the oldest seventy-four. Working on the assumption that with pupils in the same grade the younger pupil is the brighter one, there is a positive correlation with all desirable traits measured in this study. The correlation with all school marks for the six years before coming to Speyer School is .57 (Pearson formula); with the composite of the eleven tests, 1916, it is .21, while with the composite of 1917, it is .23; with the school marks the first year at Speyer it is .34, and with the ranking of the teachers at the end of one year, .30. There is, of course, nothing startling about these correlations. It is to be expected that the brighter a pupil, the quicker he will get to junior high school or to any other desirable objective point in his school career. While age was not considered in the original grouping of the pupils in this study, it is evident now that it could have been used with possibly some profit. Since the correlation of age with the composite of the eleven tests—1916, .21, and .23, 1917—is lower than the correlation of age with previous school marks, .57, or with marks at Speyer, .34, or with teachers' ranking at the end of one year, it seems to follow that these tests are a less effective measure of mental ability than the judgments of teachers, or it calls in question T. L. Kelley's statement that "the use, as a measure of intelligence, of the age at which a pupil reaches a certain grade gives the brighter pupil but a part of the credit due him." Otherwise, why is the correlation between youth and the tests not as high as that between youth and school marks?

6. SCHOOL MARKS

If school marks represented some definite achievement the difficulty of knowing their worth would be greatly simplified. While it can be shown by everyone who cares to try the experiment that in marking any paper there is much less agreement than is desirable in the subjective judgments of a group even of experts, yet aside from objective measurements school marks are one of the best standards we have of mental ability. In this study there has been an attempt to refine the value of marks. The teachers of these pupils have held that if it is desirable to hold a fast-moving class up to *x* quality in efficiency, it is likewise desirable to bring a slow-moving class as near as possible to the same degree of efficiency.

It seemed, therefore, since the groups move at different speeds, that a mark of B in Group 1 was not the same thing, when quantity as well as quality of work done is considered, as the same mark in Group 6. To equalize the difference caused by the more rapid work of the faster groups, the school marks were turned into figures, and the percentage of the original mark represented by the fraction of a school year that a class was ahead or behind the expected speed—usually the work of the middle groups—was added to or subtracted from the mark. However, this treatment did not disturb greatly the ranking made by the unweighted marks. The correlation between the weighted and unweighted marks was .94.

Since all the teachers at Speyer taught together, were under the same supervision, and at teachers' meetings frequently discussed the meaning and distribution of marks, made up and put into use a form of report card of their own, it was not difficult to turn the letters given as school marks into figures. In considering the marks made by the seventy-four pupils in the six years in many public schools under a great number of different teachers before coming to Speyer, it was but natural that the difficulty of turning the different school marks into figures was greater. To overcome this difficulty, various teachers from different public schools in the neighborhood of Speyer were asked to translate into figures the letters used in marking. The median value of a letter as found by this investigation was

used in translating the marks of the first six years into figures. With this preliminary work done, the marks made the year previous to coming to Speyer were correlated with the academic marks made the first year at Speyer. The correlation was .42. When all marks that the pupils had made before coming to Speyer were correlated with all marks made in academic subjects during their first year there, the correlation was raised to .49. However, the correlation between the composite of the eleven tests given for the purpose of classifying the pupils when they entered the school and the marks in academic subjects made by these pupils during this first year in the school, was higher still. This correlation was .57. The tests, then, were a better means of prognosis for these pupils when they entered Speyer than were all their previous school marks. The same superiority of the tests over the marks for the first six years is shown when these marks and the 1916 tests are compared with the teachers' ranking after teaching the 'pupils one year'. The correlation between the marks for the six years before coming to Speyer and the ranking by four teachers after teaching the pupils one year was .50, while that between the 1916 tests and the teachers' ranking was .66.

7. TRANSFERS

A still more practical evaluation of the accuracy of the organization into homogeneous groups can be arrived at by considering the transfers made by the teachers during one year. It will be recalled that a group or class, due to the size of the class rooms, contained only twenty-five pupils in the beginning, and that it was necessary to maintain about that size class; also that when the teachers of a pupil considered that he was in too slow or too rapid a group, they transferred him to a slower or faster class. By the final placing of the pupils as represented by the ranking of the teachers at the end of one year, ten pupils were transferred twenty-five places or more from that assigned them by the eleven tests given one year previously. Had the classes or groups contained thirty pupils instead of twenty-five, this would have been still smaller than the usual classes in New York City junior high schools or intermediate schools. With

classes of thirty there would have been only five displacements.
If it is kept in mind that the correlation between school marks
and the composite teachers' ranking is .90, it is evident that
these teachers did not as a rule make any great distinction be-
tween school marks and general mental ability.   For example,
pupil number 4 as marked by the original tests was out of school
for three weeks on account of illness, and pupil number 7 was
out for a month with an operation for appendicitis.   These two
pupils were placed much lower than ranks 4 and 7 by the teach-
ers at the end of one year.   The purpose is not to dwell on
what might have been had illnesses been unknown and teachers
omniscient, but to point out (1) that the tests foretold more
clearly than did all previous school marks the academic success
that the pupils would make at Speyer; (2) that, by the final
ranking of the teachers, there were only ten displacements of
twenty-five places or more; and (3) that had the school classes
or groups contained thirty pupils there would have been only
five displacements.

8. TEACHERS' RANKINGS

It is not supposed that the judgment of a teacher, even after
teaching a pupil for one year, is one hundred per cent perfect.
If teachers' judgments were absolutely accurate, there would
be perfect correlation between teachers 1, 2, 3, and 4.   (Table
C)   Instead of perfect correlation, however, the correlations
between teachers' rankings vary from .87 to .45, while the aver-
age of the correlations of teachers 1, 2, 3, and 4 with the other
three is .69, .67, .53, and .55.   Since teacher number 1 has the
highest average correlation, .69, with the other three teachers
and also the highest correlation with the tests, .67, for 1916 and
.73 for 1917, there is evidently some ground for the belief that
this teacher's judgment of the general mental ability of the
pupils is more accurate than that of any other of the four teach-
ers.   In the same way, since teacher number 2 has an average
correlation of .65 with the others and a correlation of .65 with
each of the composites of the tests, this teacher can be justly
ranked as second.   If the rankings of the pupils by teachers 1
and 2 be combined and reranked on the basis of the totals, the

correlation of the combined judgments with the composite of
the 1916 tests will be .69 and with the 1917 tests, .72. The com-
bined judgments of teachers 1, 2, 3 have a correlation with the
composite of its 1917 tests of .73. However, when teacher num-
ber 4 is introduced, the correlation falls to .68. This combina-

TABLE C

THE RELATION OF ONE TEACHER'S RANKING TO THAT OF ANOTHER

	Teacher No. 1	Teacher No. 2	Teacher No. 3	Teacher No. 4	Composite of Tests, 1916	Composite of Tests, 1917
Teacher No. 1..........		.87	.58	.61	.67	.73
Teacher No. 2..........	.87		.56	.58	.65	.65
Teacher No. 3..........	.58	.56		.45	.47	.54
Teacher No. 4..........	.61	.58	.45		.45	.32
Composite of Tests, 1916	.67	.65	.47	.45		
Composite of Tests, 1917	.73	.65	.54	.32		

tion of teachers' rankings and the relation of the resultant
ranking with the ranking by the tests, is emphasized here to
show that teachers' judgments do vary and that if the teachers
had been selected, slightly higher correlations would have been
found. It will be remembered that the judgments used were
those of all teachers who had taught all of these boys in academic
subjects. While the stressing of this point is of little impor-
tance, the fact is to be noted that the correlation of the com-
posite of the teachers' judgments of pupils with the composite
of the eleven tests of 1916 is .66, and with the composite of the
1917 tests the correlation is .68; also the fact that teachers 1, 2,
3, and 4 correlate with the 1916 tests, .67, .65, .47, and .45, and
with the 1917 tests, .73, .65, .54, and .32. These facts make it
clear that the teachers individually really agree with the rank-
ing by tests as well as they agree with each other. This addi-
tional point should be noted,—that the correlation of the com-
posite of the four teachers' judgments with the composite of
the tests is decidedly higher than the average of their corre-
lations with each other.

9.   The First Question Answered

The first question proposed in this study was: ''To what extent is the attempt at educational prognosis made on the basis of eleven certain standardized educational and psychological tests in agreement with the judgment of four teachers after teaching the pupils tested for one year?'' This question has now been answered. With temporary illnesses and all the varying interests that come in one year to the boy of twelve or thirteen, in the opinion of four teachers at the end of one year only ten pupils, on the basis of twenty-five in a class, had been originally placed in too high or too low a class, and, on the basis of thirty in a class, only five. Further, the success, as represented by school marks, of seventy-four boys just a very little better than the median pupil in twenty-four 6B class rooms of five New York City public schools, was more accurately predicted by eleven standardized tests than by all the pupil's previous marks combined.

CHAPTER III

FOR THE PURPOSE OF EDUCATIONAL PROGNOSIS, WHAT TESTS ARE OF MOST VALUE? STANDARDS

WITH the first problem concerning the possibility of making an educational prognosis by means of standardized tests answered, in so far as the data of this study permit, there arises the question of evaluating these tests for the purpose set forth in this problem. Which of these tests, how many tests, and what combination of them must the practical administrator give in order to arrive at as good results or even better than those reached in this study? It is not maintained that eleven tests are a sufficient number; the more measures of equal value, the better. Neither is it maintained that tests that can be given to a whole group at one time are more accurate in making a diagnosis of the pupil than are tests which can be given to only one subject at a time. The complete study of a single individual would occupy a lifetime. However, the problem here is to select from the eleven tests used those tests which, with due regard to economy of the pupils' time and ease in scoring, the administrator can use in organizing, for purposes of instruction, the entering classes in the junior and senior high schools.

Seven standards are proposed for evaluating these tests:
1. The correlation of a test with itself, or with a similar test, repeated with the same pupils one year after the first test is given.
2. The correlation of each test with the composite of the eleven tests.
3. The correlation of each test with each of the other ten tests separately.
4. The correlation of each test with the judgments of four teachers after teaching the pupils for one year.

5. The correlation of each test with all the school marks the pupil made during his school life before he reached the junior high school.

6. The correlation of each test with the school marks in all academic subjects during the first year in junior high school.

7. The correlation of each test with the age of the pupil.

Since all the tests, either the same or similar ones, were repeated in the present study, all of the standards proposed above, with the exception of number 1, have been worked out twice. In addition to this repetition, in order to correct each test for attenuation, each 1916 test was correlated with each 1917 test. At this point, when each of these seventy-four pupils had participated in several hundred correlations, certain tests were selected as being the best for the purposes of this study. These tests, as will be seen, are further correlated and combined so that the administrator may know the degree of efficiency he may expect for the number of minutes invested in measuring the pupil.

It is not maintained, of course, that all of these standards are of equal value. For example, it is possible that standard 1 may be of slight value. The justification of correlating each test with age may be called in question. However, in every case in this study, which includes only 6B pupils, when the youngest pupil has been ranked 1, and the oldest 74, there has been a positive correlation between youth and the ranking by tests, by school marks, or by teachers' ranking. Standard 3 probably should not be considered of too great value if it were too much opposed to standard 2. The question may be raised as to the reason for not making more use of the coefficients that result from the correction for attenuation. The value of the elimination of chance error as represented by this process has not been overlooked, but it is believed for the purposes of the present study that it is safer to depend on what the administrator in using tests will have to depend on—the raw coefficients. One further question is considered. When the best one, two, or three or half dozen tests have been selected, it is possible to find out what would have been the result if these tests and these only, instead of the eleven tests, had been used in making the original prognosis.

1. The Tests Repeated

As has been pointed out, the tests given in February, 1916, were repeated one year later, February, 1917. Of these 1917 tests, it will be recalled that one test, Reading Alpha 2: Part II, was the same test, that the Woody tests of 1917 were the same as the 1916, but that only about half as many problems were used. The other eight tests while not identical were, as is pointed out on pages 9 and 10, similar. In correlating the score of each 1916 test with its corresponding test in 1917 (see Tables Y and Z, pages 51-53) the Spearman method was used. According to the formula

$$\rho = 1 - \frac{6\Sigma D^2}{n(n^2-1)}$$

ρ is the measure of the correlation, $n =$ the number of paired related measures, $D =$ the difference in rank of the subject in the measures correlated, and $\Sigma D^2 =$ the sum of the differences squared, or, to put it more concisely, $\Sigma D^2 =$ ''the sum of the squares of the differences between the two numbers denoting the relative positions of the two related measures in their respective series.'' Since it is necessary to have the coefficient in terms of r, the Pearson formula, in order to employ the formula for correction for attenuation, the coefficients worked out by the Spearman method have been in every case transmuted, according to the table [1] for inferring the value of r from any given value of ρ, into coefficients in terms of the Pearson formula. The reliability of the coefficients derived is, of course, dependent on the number of cases. In this study it will be recalled that there are seventy-four pupils or cases. The P.E., then, as the ''median of the differences between the separate measures and their central tendency,'' shows the measure of reliability. By the formula,

$$\text{P.E.} = \frac{.6745\,(1-r^2)}{\sqrt{n}}$$

$n =$ the number of cases and $r =$ the coefficient of correlation.

[1] Thorndike, E. L., *Mental and Social Measurements*, Table 36, p. 168, 1913 edition.

Therefore:

The Coefficient of Correlation of .10 has a Probable Error of .08
"   "   "   "   "   .20   "   "   "   "   "   .07
"   "   "   "   "   .30   "   "   "   "   "   .07
"   "   "   "   "   .40   "   "   "   "   "   .07
"   "   "   "   "   .50   "   "   "   "   "   .06
"   "   "   "   "   .60   "   "   "   "   "   .05
"   "   "   "   "   .70   "   "   "   "   "   .04
"   "   "   "   "   .80   "   "   "   "   "   .03
"   "   "   "   "   .90   "   "   "   "   "   .01

In order to be sure to have the same response from two situations, the stimulus and the situations must be exactly the same. As applied to the tests repeated here, ''the same response'' would mean perfect correlation between the 1916 and 1917 tests. The correlation, however, is by no means perfect. Some of the factors that keep one from expecting too high a correlation between the two tests demand consideration. The tests repeated, as we have seen in the portion of this study devoted to a discussion of the tests given, while always similar, were not always the same ones. Trabue B and C undoubtedly measure the same abilities as Trabue J and K, yet the tests are by no means of the same difficulty. While the directions given to pupils, when Reading Alpha 2: Part II was given in 1917 were the same as those given for this test in 1916, no one can be sure that the mental set of the pupils was the same. In fact, one can be sure that it was not the same. In 1916 these pupils were not accustomed to taking tests of this kind, while a year later they counted it a dull day that they did not have a chance to measure themselves by some objective standard. Then, too, while the pupils tested were the same ones, a whole year had elapsed between the tests and their repetition. Individual differences that existed in 1916 had, as a result of opportunity for practice and especially practice in groups that stimulated one to progress at one's optimum speed, increased rather than equalized these differences. While it is not the object of this study to stress the enormous changes that took place in one year in boys twelve and thirteen years of age, yet, in considering the correlation between the 1916 and 1917 tests, it is necessary to recognize that great changes at this period are possible. There are in all probability errors other than those which correction

for attenuation can eliminate. This point, however, will be discussed later. The fact that the year elapsing between tests and their repetition brought changes in the pupils tested, that the mental set of these pupils was different, and that the tests were not in all cases exactly the same, must be taken into account in considering the correlations presented in Table D.

TABLE D

THE CORRELATION OF A TEST WITH ITSELF, OR WITH A SIMILAR TEST, WHEN REPEATED WITH THE SAME PUPILS AFTER ONE YEAR

					Rank
Visual Vocabulary	1916 with	Visual Vocabulary	1917	.56	(1)
Reading Alpha 2: Pt. II	" "	Reading Alpha 2: Pt. II	"	.52	(2.5)
Composition	" "	Composition	"	.32	(10)
Spelling	" "	Spelling	"	.52	(2.5)
Trabue B	" "	Trabue J	"	.36	(7.5)
Trabue B	" "	Trabue K	"	.22	
Trabue C	" "	Trabue J	"	.29	
Trabue C	" "	Trabue K	"	.30	(11)
Trabue B and C combined	" "	Trabue J and K combined	"	.41	
Woody Multiplication	" "	Woody Multiplication	"	.43	(5)
Woody Division	" "	Woody Division	"	.38	(6)
Opposites	" "	Opposites	"	.44	(4)
Easy Directions	" "	Easy Directions	"	.34	(9)
Mixed Relations	" "	Mixed Relations	"	.36	(7.5)
Composite of All Tests	" "	Composite of All Tests	"	.79	

It is not possible to compare accurately these raw coefficients of correlation with the work of any other investigators of whom the writer knows, for in other cases either these tests have not been given, or they have not been given a year apart, or to boys of twelve and thirteen. As they stand here, if Trabue B is paired with J, and C with K, the tests rank, in degree of correlation, according to the figures in parentheses following the coefficients of correlation in the total.

The worth of this standard is problematical. Should a test repeated with the same pupils after one year have a high correlation with itself? For example, English Composition had not received nearly the emphasis in the first six grades of the public school that it did in the year between these tests. If each pupil had improved, say twenty per cent, during the year, the ranking in composition by the 1916 test would not have been disturbed; but the improvement in each pupil's case was not, in comparison with the other pupils, a certain per cent of his orig-

inal ability.   If it had been, the correlation when corrected for attenuation would have been more nearly perfect.   Such, however, as will be seen later, is not the case.   It is possible that if this subject, English Composition, had received still more emphasis during the time between the tests, the correlation between the two tests would have been still lower.   While the pupils had had much practice in composition, the opposite is true regarding the type of test represented by Visual Vocabulary.   To be sure, the pupils had been learning new words, but they had had no practice in writing F under a word that means a flower, or T under a word indicating something about time, yet the correlation in the case of Composition is .32, and in Visual Vocabulary, .56.   Spelling had received great attention during the first six years and the study of this subject was continued during the year between the tests, yet the correlation is but .52. Differences in tests, in mental set, in physical condition, in the lapse of one year, furnish some of the explanations of the low correlations, and at the same time call in question the worth of this standard in determining the value of a test for purposes of prognosis.

2.   THE CORRELATION OF EACH TEST WITH THE COMPOSITE

The second standard proposed is the correlation of each test with the composite of the eleven tests.   The method of doing this has already been explained.   Exactly to what extent the various tests measure different mental traits it is not yet possible to say.   However, since all tests used have been found to correlate positively with desirable mental abilities for academic work, it seems fair to assume that all the tests as combined in the composite give a more accurate evaluation of the pupil's general mental ability than does any one of these tests singly. Therefore, it follows that the correlation of each test with the composite furnishes some means of evaluating the relative merits of the different tests.   Since the tests have been repeated, a possible check on the value of a test as determined by its correlation with the composite, is furnished.   If the possible reasons for causing the high or the low correlation of a test with itself when repeated, are held in mind, the comparison of the corre-

lation of each test in 1916 and in 1917 with its composite will be of value.

It will be noted in Table E that with the exception of Opposites and Easy Directions, the tests occupy somewhat the same relative positions in 1916 and in 1917. Visual Vocabulary, for example, which ranked 1 in 1916, becomes 3 in 1917, Reading changes from rank 3 in 1916 to rank 4 in 1917, and so on. If the ranks for each test for each year are added and these totals ranked, and if the correlation of each test with its composite in 1916 be added to its correlation with its composite in 1917, and these totals ranked, and then these two totals added and ranked, the tests will stand in the relative positions expressed by the column of figures in parentheses in Table E.

TABLE E

CORRELATION OF EACH TEST WITH ITS COMPOSITE

		1916		1917	Rank
Visual Vocabulary		.73		.69	(1)
Reading		.63		.67	(2)
Composition		.51		.50	(8)
Spelling		.53		.54	(7.5)
Trabue	B	.45	J	.63	(7.5)
Trabue	C	.59	K	.65	(3)
Trabue B and C		.65	J & K	.76	
Woody Multiplication		.26		.36	(9)
Woody Division		.26		.30	(10)
Opposites		.49		.70	(4)
Easy Directions		.58		.52	(5.5)
Mixed Relations		.55		.54	(5.5)

It will be observed that while it is possible and probably just, in considering the Trabue tests, to pair B with J and C with K, yet it might have been wiser in the beginning to have combined B and C and J and K. This doubling the length of the test makes these completion tests take a relatively higher position. It is apparent from the correlations, according to the standard considered here—the degree of correlation of a test with its composite,—that Visual Vocabulary, Reading, Opposites, and Trabue Completion are the four tests of greatest value for purposes of educational prognosis.

There is, however, a source of error in all these correlations which makes them higher than they should be; this is especially true of the combination made of the Trabue tests. The compos-

ite is composed of eleven separate tests, and when any one of
these tests is correlated with this composite, it is, to a certain
extent, correlated with itself.   Since the Trabue tests B and C,
likewise J and K, enter into the make-up of the composite as two
separate tests, they have, when combined, a double interest in
the composite.   Statistically, it would be equally just to com-
bine Visual Vocabulary and Reading.   This combination has a
correlation with the composite of .76 in 1916 and .74 in 1917.
Since eleven tests enter into the composition of the composite,
each test would seem to have an interest of one-eleventh, and
when two tests are combined after the composite has been made
up, as was done with Trabue B and C and also with J and K,
such a combination would have an interest of two-elevenths in
this composite.   Investigators as a rule have not made any cor-
rection for this correlation of a test with the composite of which
it is a part.   Manifestly, however, such a correction is of value.
One of the ways of making this correction that suggests itself
is to make a composite of ten tests and find the correlation be-
tween this composite and the eleventh test.   This method is not
only exceedingly laborious but evidently partly unjust.   The
test withdrawn from the composite in order to correlate it with
the other ten tests, measures some phase of general mental abil-
ity, and with this test withdrawn the composite is proportion-
ately less perfect.   However, this method has been followed.
Each of the eleven tests has been correlated with the composite
of the other ten.

The results of the correlation of every test with the composite
of the other ten tests, as presented in Table F, show that the
correlations as presented in Table E have been reduced about
.16 in 1916 and about .13 in 1917.   The reductions, when this
method is used, for each individual test as shown by the figures
in parentheses in Table F, vary in 1916 from .13 to .21, and in
1917 from .07 to .18.   It will be noted also that this reduction
is not a fixed percentage of the original correlation, the correla-
tion of the test with the composite of the eleven tests, but that
the amount of reduction has a slight tendency to be larger when
the original correlation is smaller.   This correction for the cor-
relation of a test with itself does not materially affect the order

of tests as ranked in Table E. However, it does raise again the question as to the extent of common elements in the various tests.

TABLE F

CORRELATION OF EACH TEST WITH THE COMPOSITE OF ALL
TESTS EXCEPT ITSELF

		1916			1917	
Visual Vocabulary		.60	(.13)		.56	(.13)
Reading		.47	(.16)		.56	(.11)
Composition		.35	(.16)		.36	(.14)
Spelling		.40	(.13)		.37	(.17)
Trabue .... B		.29	(.16)	J	.50	(.13)
Trabue C		.45	(.14)	K	.53	(.12)
Woody Multiplication		.09	(.17)		.22	(.14)
Woody Division		.06	(.20)		.17	(.13)
Opposites		.28	(.21)		.63	(.07)
Easy Directions		.42	(.16)		.41	(.18)
Mixed Relations		.37	(.18)		.41	(.13)

3. THE CORRELATION OF EACH TEST WITH EVERY OTHER TEST

The correlation of each test with every other test proceeds on the assumption that each of these unweighted tests is equal to any other test. By the combination of the seven standards set up for evaluating a test, this, as is pointed out later, is found to be untrue. Such value as the correlation of each test with every other test has, is not to be neglected; but if this standard is in conflict with the corrected correlation of each test with its composite as presented in Standard 2, its value would certainly be questionable.

By Tables G and H it is possible in addition to knowing the correlation of every test with every other test and the relations between these correlations for the 1916 and 1917 tests, to know, also, how this standard of the average correlation of each test with the ten other tests composing its composite, when the 1916 and 1917 averages are combined, compares with the second standard set up—the correlation of each test with its composite. In combining the 1916 and 1917 tests by ranking the combined totals of the ranks arrived at, first, by ranking each test according to the total correlations with every other test in both 1916 and 1917, and, second, by ranking the totals given by adding the ranks of each test in 1916 and 1917, it is found that the tests according to this standard stand in the following order with

the highest first: Visual Vocabulary, Reading, Opposites, Trabue C-K, Easy Directions, Mixed Relations, Spelling, Trabue B-J, Composition, Woody Multiplication, and Woody Division. By comparing this ranking of tests just given in the order of their importance for educational prognosis with the corresponding order arrived at by the correlation of each test with its composite, Table E, it will be seen that the order of the tests is almost the same.

TABLE G
EACH 1916 TEST WITH EVERY OTHER 1916 TEST

	Visual Vocab.	Reading	Composition	Spelling	Trabue B	Trabue C	Woody Mult.	Woody Division	Opposites	Easy Directions	Mixed Relations
Visual Vocabulary		.49	.41	.29	.23	.43	—.08	.05	.34	.46	.47
Reading	.49		.25	.24	.33	.36	.00	.01	.10	.46	.32
Composition	.41	.25		.15	.09	.36	.17	.03	.24	.21	.08
Spelling	.29	.24	.15		.20	.15	.27	.26	.17	.29	.10
Trabue B	.23	.33	.09	.20		.29	.13	—.04	.25	.07	.10
Trabue C	.43	.36	.36	.15	.29		.04	—.06	.32	.23	.46
Woody Multiplication ...	—.08	.00	.17	.27	.13	.04		.40	.03	—.25	.03
Woody Division	.05	.01	.03	.26	—.04	—.06	.40		—.09	.08	—.05
Opposites	.34	.10	.24	.17	.25	.32	.03	—.09		.30	.22
Easy Directions	.46	.46	.21	.29	.07	.23	—.25	.08	.30		.31
Mixed Relations	.47	.32	.08	.10	.10	.46	—.03	—.05	.22	.31	

TABLE H
EACH 1917 TEST WITH EVERY OTHER 1917 TEST

	Visual Vocab.	Reading	Composition	Spelling	Trabue J	Trabue K	Woody Mult.	Woody Division	Opposites	Easy Directions	Mixed Relations
Visual Vocabulary		.61	.26	.31	.32	.50	.24	—.02	.50	.33	.32
Reading	.61		.31	.23	.28	.43	.23	.13	.51	.26	.29
Composition	.26	.31		.18	.27	.27	.04	.27	.41	.10	.05
Spelling	.31	.23	.18		.25	.24	.44	.12	.43	.12	.10
Trabue J	.32	.28	.27	.25		.42	.00	.15	.45	.43	.31
Trabue K	.50	.43	.27	.24	.42		.14	.10	.50	.27	.23
Woody Multiplication ...	.24	.23	.04	.44	.00	.14		.18	.38	—.08	.21
Woody Division	—.02	.13	.27	.12	.15	.10	.18		.36	.19	.23
Opposites	.50	.51	.41	.43	.45	.50	.38	.36		.52	.50
Easy Directions	.33	.26	.10	.12	.43	.27	—.08	.19	.52		.36
Mixed Relations	.32	.29	.05	.10	.31	.23	.21	.23	.50	.36	

The determination of the value of a test for educational prognosis, however, is not the whole question here; if it were, the whole problem would be greatly simplified. The question is not only what tests are best for the purposes of prognosis, but what combination of tests is desirable. If every test in Tables G and H had a correlation of $+1$. with every other test, there would be no need of giving eleven tests; one would do as well as all combined. In such a case it would be evident that all tests had measured the same function. Visual Vocabulary, Reading, and Completion Tests tend to measure at least closely related functions, as is shown by their correlations with each other. A test that has shown positive correlation with desirable traits and has a low correlation with every other test, evidently measures a function not measured by these other tests. This accounts for the negative correlation of the Woody tests in Tables G and H. In measuring a group it is, of course, desirable to measure as many traits as possible. Thus a test that measures traits not closely related to those measured by the other tests will have a low correlation with the other tests, and at the same time be the test that should be included in the combination of tests used for the purpose outlined in this study. On this basis the test with the lowest correlation in Tables G and H has been ranked one, and the test with the highest correlation, eleven.

4. The Correlation of Each Test with the Judgment of Four Teachers

As has been pointed out, the criterion of prognosis in this experiment had to rest in the teachers' judgments. It is not believed that these judgments are always correct. In fact, one can be sure that some of them at least are incorrect, for the average correlations of each of the four teachers' judgments with those of the other three, as has been pointed out, are .69, .67, .53, and .55 instead of $+1$., as they would be if the teachers were omniscient. However, such virtue as lies in this study in spite of such imperfections as may exist in material, method, or individual judgments, is due largely to the fact that it is a study of a practical working experiment. Since such is the case, teacher-

judgments are accepted as they are without theorizing as to what they might be.

It will be recalled that the correlation of the Teachers' Ranking with the composite of the 1916 tests is .66 and with that of the 1917 tests, .68. It is not to be expected that any one test will reach as high a correlation with the Teachers' Ranking as the composite of all the tests unless some of these tests are worthless, for the purpose considered here, or worse than worthless, or that some of the tests are of such a compound of many tests as to measure a very great number of mental traits that correlate positively with those mental traits that make for academic success. Each of the eleven tests, for both 1916 and 1917, as shown in Table J, has been correlated with Teachers' Rankings. By inspection, those tests which correlate highly with Teachers' Ranking can be easily picked out. However, to arrive at a definite statement, some statistical method is necessary. By ranking each test for 1916 and for 1917 and ranking the totals of these tests, or by ranking the tests by the average of the correlation of each test in 1916 and 1917, there is very little changing of the relative position and there is no change in that of the six highest correlations. By adding the rankings by each method and ranking the totals, the tests stand in the order indicated by the figures in parentheses: (3), (2), (4), (1), etc. It will be observed that in this ranking Trabue B and C have been combined and also J and K. Thus there are only ten tests. However, instead of making this combination, if C had been paired with K, and B with J, with the resultant ranking as shown by the figures (3), (1.5), (4), etc., the method of ranking being the one just explained, the only difference so far as these Completion Tests are concerned, is that C-J takes the place of the longer tests. The point is often rightly urged that lengthening a test tends to raise its correlation. Lengthening a test, however, means that it takes more time for the subjects to take it, and likewise a longer time for the administrator to score it. For theoretic purposes, time is not of so great value; but for practical use, if C-K will give as satisfactory a result as will B and C and J and K, then according to the standard now being considered for evaluating a test, C-K is to be preferred.

TABLE J

CORRELATION OF TESTS WITH THE RANKING BY FOUR TEACHERS

				Rank Trabue B and C combined, J and K combined	Rank Trabue B paired with J, and C with K
Visual Vocabulary	.44		.43	(3)	(3)
Reading	.47		.47	(2)	(1.5)
Composition	.37		.49	(4)	(4)
Spelling	.37		.60	(1)	(1.5)
Trabue B	.18	J	.40		(9)
Trabue C	.38	K	.24		(6)
Trabue B and C	.37	J & K	.36	(6)	
Woody Multiplication	.25		.35	(7.5)	(8)
Woody Division	.20		.35	(9)	(10)
Opposites	.36		.50	(5)	(5)
Easy Directions	.35		.27	(7.5)	(7)
Mixed Relations	.22		.18	(10)	(11)

It may be recalled that in Table B, where the rankings by fourteen teachers, each one ranking his or her own pupils, were compared with the ranking by the four tests, Spelling and Arithmetic correlated about twice as high with the Teachers' Rankings as did Composition and the B and C Completion Tests. An inspection of Table J shows that while the Arithmetic tests do not rank so high as in Table B, Spelling leads the list: Right or wrong, the ability that enables a pupil to spell well plays an important part in forming a teacher's conception of mental ability. In contrast to this important place maintained by spelling, the Arithmetic tests are here among those that have the lowest correlations with Teachers' Rankings. Plainly the order of the first half-dozen tests according to the standard now under consideration is: Spelling, Reading, Visual Vocabulary, Composition, Opposites, and the Completion Tests.

5. RELATION OF EACH TEST TO SCHOOL MARKS DURING THE FIRST YEAR OF THE JUNIOR HIGH SCHOOL

Since the correlation of teachers' judgments and the school marks is so high, .90, it seems evident that those boys who do their school work well are, in the opinion of the teachers, the abler mentally. No such close relation is found between the tests and school marks. This correlation for 1916 is .57, and

for 1917, .55. Since the two standards, school marks and teachers' judgments, are both subjective, and since the four teachers whose combined judgments determine the rankings of the pupils, gave about half the marks, a close correlation between these two standards is to be expected. In all probability, a wider range of mental traits is represented in teachers' marks than is measured in any one of the eleven tests. Thus, while the correlation between the school marks in the first year of the junior high school and the composite of the 1916 tests is .57, and the composite of the 1917 tests is .55, the correlation of school marks of the first six years with the 1916 tests is .29, and with the composite of the 1917 tests, .32. As can be seen in Table K, the correlations of the individual tests with the school marks during the first year of the junior high school range from .43 to .16 in 1916, median .29, and in 1917, from .56 to .09, with a median of .34. As has been pointed out, teachers' judgments and school marks are often variable; yet, outside of objective measurements, they are the best measures of general intelligence that we have. It follows, therefore, that school marks should receive some consideration in evaluating a test.

In studying Table K, it will be noted, when all tests are considered, that the average correlation of the 1917 tests is

TABLE K

CORRELATION OF EACH TEST, 1916 AND 1917, WITH SCHOOL MARKS IN ACADEMIC SUBJECTS DURING THE FIRST YEAR OF JUNIOR HIGH SCHOOL

	1916		1917	*Rank*
Visual Vocabulary	.34		.32	(5.5)
Reading	.43		.37	(2)
Composition	.32		.38	(3)
Spelling	.32		.56	(1)
Trabue B	.16	J	.29	(9.5)
Trabue C	.31	K	.09	(9.5)
Woody Multiplication	.28		.39	(5.5)
Woody Division	.27		.34	(7)
Opposites	.26		.47	(4)
Easy Directions	.29		.18	(8)
Mixed Relations	.17		.18	(11)

slightly higher than that of the 1916 tests. There is at the same time great variation in the ranking of the tests: Visual Vocabulary, which ranked 2 in 1916, is ranked 7 in 1917, Oppo-

sites has jumped from 9 to 2, and Reading with a change of only .06 in correlation, has dropped from first to fifth place. By the method of ranking already explained, the tests stand, when school marks are considered as a standard for evaluation, in the order indicated by the figures in parentheses in the right-hand column. As will be noted, Spelling holds first place, as it did with Teachers' Rankings. Reading is second, Composition third, Opposites fourth, with Visual Vocabulary and Woody Multiplication tied for the next position. While the order is varied, and, with the exception of Woody Multiplication taking the place often held by the Completion Tests, the first half-dozen tests here are the same as those selected by the preceding standards.

The tests used were selected, as has been pointed out, because in previous experiments they had had positive correlations with desirable traits as shown by academic success, and because it was believed that this general mental ability under right direction would express itself in the school work. Hence a positive relation was expected between the tests and school marks. If the school work to be done had been other than that of an academic junior high school, it is conceivable that some other or some additional tests might have been selected.

6. THE CORRELATION OF EACH TEST WITH ALL SCHOOL MARKS
MADE DURING THE FIRST SIX YEARS

Perhaps no absolutely positive statement concerning the pupil's mastery of the "tool subjects" in the first six grades, as usually taught, and his general mental ability, can be made. One is certainly justified, it would seem, in believing that the pupil with mental ability would master such subjects as the four fundamentals in arithmetic and thus rank high according to the marks that he received as a result of doing this work well. When it is recognized that in this study the correlation of all school marks made prior to entering the junior high school with all marks made during the first year after entering, is .49, while that of the composite of the 1916 tests is .57, and, at the same time, that the correlation of the marks for these first six years

with the Teachers' Ranking at the end of the first year of the junior high school is .50, while that of the 1916 tests is .66, it is evident that for predicting academic success the tests were superior to the sum of all previous marks. The recognition of this fact calls in question the value of this standard of previous school marks for evaluating a test, and especially marks given in so many grades of so many schools by so many teachers.

Since the work of the first six grades probably must be, and certainly is, on the "fundamentals," the ranking of the tests by this standard of all marks previous to the junior high school is not surprising. Thus in Table L, the tests rank, beginning with the highest: Spelling, Woody Division, Composition, Trabue B-J, Woody Multiplication, with Reading and Opposites tied for the sixth place, followed by Visual Vocabulary, Mixed Relations, Easy Directions, and Trabue C-K. Plainly there is, with the exception of Spelling, a marked reversal in the order of the tests from what has been found in the other standards.

TABLE L

THE CORRELATION OF EACH TEST FOR 1916 AND 1917 WITH ALL SCHOOL MARKS BELOW THE JUNIOR HIGH SCHOOL

	1916		1917	Rank
Visual Vocabulary	.15		.24	(8)
Reading	.22		.20	(6.5)
Composition	.21		.29	(3)
Spelling	.40		.36	(1)
Trabue B	.16	J	.33	(4)
Trabue C	.13	K	.07	(11)
Woody Multiplication	.29		.18	(5)
Woody Division	.19		.37	(2)
Opposites	.16		.28	(6.5)
Easy Directions	.10		.15	(10)
Mixed Relations	—.02		.24	(9)

The high position of Spelling and Arithmetic can be easily understood—these subjects had received emphasis in the first six grades. Certainly the brighter pupils should master the fundamentals of arithmetic better than the dull ones, and likewise spell better. However, does the habit of making a fixed response to a situation, instead of freeing the mind for other things, interfere for the time during which the habit is being fixed, with meeting entirely new situations? In 1916 and again in 1917, the Multiplication tests correlated negatively with Easy Direc-

tions. Such tests as Reading and Visual Vocabulary which, according to other standards so far considered, rank high, are here in the second division. By this standard of marks for the first six years, those tests involving fixed responses rank much higher than those involving new situations. Aside from these observations, since the school marks under consideration have a correlation of .49 with all marks in the first year of the junior high school, of .50 with the rankings by four teachers, and since all of these marks had a correlation of only .29 with a composite of the 1916 tests and an average correlation with all the tests of only .18, it is not believed that this standard is of much value in determining the worth of a test.

7. THE CORRELATION OF EACH TEST WITH THE AGE OF THE PUPIL

Since retardation and at least comparative acceleration play some part in every school system, it is to be expected that within a grade the younger pupils have the greater mental ability. Hence, as has been pointed out, the youngest pupil has been ranked 1, and the oldest, 74. Yet from the data presented in Table M, it is seen that there is a very low correlation between youth and the tests. The average of the correlations of all tests for 1916 with youth is .13, and for 1917, .17, while the correlation of the composite of all the tests for these years, 1916 and 1917, with youth is .21 and .23. The bright young pupils evidently attracted the favorable attention of the various teachers during the first six years, for the correlation between the marks for the first six years and youth is .57. However, these comparatively accelerated pupils did not succeed quite so well, as judged by school marks, during the first year of the junior high school. Here the correlation between school marks and youth is not .57 but .34, and the correlation with the Teachers' Ranking at the end of one year is .04 lower.

In analyzing Table M, it will be noted that the tests selected by the standard of youth are, first of all, those preferred by Standard 6—Arithmetic and Spelling. It should be noted, however, that Visual Vocabulary jumped from rank 9.5, 1916, to 2.5, 1917, and Opposites from 11 to 4.5. The data of this table might suggest also—since such tests as those just mentioned in-

volving situations new to these boys in 1916 were so much better
met in 1917—that comparatively these brighter, younger pupils
adjusted themselves when once the ''tool subjects'' had been
mastered, to new situations more rapidly than the older pupils.
In any case, youth within a grade does correlate positively with
the average of the 1916 and 1917 tests and with all teachers' esti-

TABLE M

CORRELATION OF EACH TEST WITH YOUTH WITHIN A SCHOOL GRADE

	1916		1917	*Rank*
Visual Vocabulary	.04		.27	(7)
Reading	.17		.15	(3)
Composition	.15		—.03	(9)
Spelling	.27		.27	(2)
Trabue B	.12	J	.25	(5)
Trabue C	.07	K	—.02	(10.5)
Woody Multiplication	.27		.29	(1)
Woody Division	.19		.08	(6)
Opposites	—.03		.26	(8)
Easy Directions	.04		.04	(10.5)
Mixed Relations	.09		.26	(4)

mates, either marks or rankings.   Therefore, youth must be of
value as a standard for evaluating a test; but since these corre-
lations are so low, it is not of great value.   Young pupils, so
far as this study is concerned, stand higher in the sympathetic
estimates of their early teachers than in the unfeeling ranking
by objective tests.

8.   SUMMARY OF ALL RAW COEFFICIENTS OF CORRELATION

Before considering the coefficients corrected for attenuation,
it will probably be convenient for the reader to have all the raw
coefficients for all tests and for all standards set up, presented
as concisely as possible.   At the expense of some necessary repe-
tition, they are brought together in Tables N and O.   Table P
presents the average of all correlations compiled from Tables
N and O.   In these tables, Trabue B, 1916, is paired with
Trabue J, 1917, and likewise C, 1916, with K, 1917.   If C is
combined with B and J with K, the correlation between B-C
and J-K is .41.

TABLE N

Correlation of Each 1916 Test With Every Other 1916 Test

	Vis. Vocab.	Reading	Composition	Spelling	Trabue B	Trabue C	Woody Mult.	Woody Div.	Opposites	Easy Direc.	Mixed Rel.	Average	Composite	Teachers' Rank	Marks First 6 Yrs.	Marks First Year Jr. H. S.	Youth	With Itself 1917
Vis. Vocab.		.49	.41	.29	.23	.43	—.08	.05	.34	.46	.47	.31	.73	.44	.15	.34	.04	.56
Reading	.49		.25	.24	.33	.36	.00	.01	.10	.46	.32	.26	.63	.47	.21	.43	.17	.52
Composition	.41	.25		.15	.09	.36	.17	.03	.24	.21	.08	.20	.51	.37	.21	.32	.15	.32
Spelling	.29	.24	.15		.20	.15	.27	.26	.17	.29	.10	.21	.53	.37	.40	.32	.27	.52
Trabue B	.23	.33	.09	.20		.29	.13	—.04	.25	.07	.10	.16	.45	.18	.16	.16	.12	.36
Trabue C	.43	.36	.36	.15	.29		.04	—.06	.32	.23	.46	.26	.59	.38	.13	.31	.07	.30
Woody Mult.	—.08	.00	.17	.27	.13	.04		.40	.03	—.25	—.03	.07	.26	.25	.29	.28	.27	.43
Woody Div.	.05	.01	.03	.26	—.04	—.06	.40		—.09	.08	—.05	.06	.26	.20	.19	.27	.19	.38
Opposites	.34	.10	.24	.17	.25	.32	.03	—.09		.30	.22	.19	.49	.36	.16	.26	—.03	.44
Easy Direc.	.46	.46	.21	.29	.07	.23	—.25	.08	.30		.31	.22	.58	.35	.10	.29	.04	.34
Mixed Rel.	.47	.32	.08	.10	.10	.46	—.03	—.05	.22	.31		.20	.55	.22	—.02	.17	.09	.36
Average	.31	.26	.20	.21	.16	.26	.07	.06	.19	.22	.20		.51	.30	.18	.29	.13	
Composite	.73	.63	.51	.53	.45	.59	.26	.26	.49	.58	.55	.51		.66	.29	.57	.21	
Teachers' Rank	.44	.47	.37	.37	.18	.38	.25	.20	.36	.35	.22	.30	.66		.50	.90	.30	
Marks First 6 Years	.15	.21	.21	.40	.16	.13	.29	.19	.16	.10	—.02	.18	.29	.50		.49	.57	
Marks First Year Jr. H. S.	.34	.43	.32	.32	.16	.31	.28	.27	.26	.29	.17	.29	.57	.90	.49		.34	
Youth	.04	.17	.15	.27	.12	.07	.27	.19	—.03	.04	.09	.13	.21	.30	.57	.34		
With Itself 1917	.56	.52	.32	.52	.36	.30	.43	.38	.44	.34	.36							

J. K.

TABLE O
CORRELATION OF EACH 1917 TEST WITH EVERY OTHER 1917 TEST

	Vis. Vocab.	Reading	Composition	Spelling	Trabue J	Trabue K	Woody Mult.	Woody Div.	Opposites	Easy Direc.	Mixed Rel.	Average	Composite	Teachers' Rank	Marks First 6 Yrs.	Marks First Year Jr. H. S.	Youth	With Itself 1916
Vis. Vocab.		.61	.26	.31	.32	.50	.24	—.02	.50	.33	.32	.34	.69	.43	.24	.32	.27	.56
Reading	.61		.31	.23	.28	.43	.23	.13	.51	.26	.29	.33	.67	.47	.20	.37	.15	.52
Composition	.26	.31		.18	.27	.27	.04	.27	.41	.10	.05	.22	.50	.49	.29	.38	—.03	.32
Spelling	.31	.23	.18		.25	.24	.44	.12	.43	.12	.10	.24	.54	.60	.36	.56	.27	.52
Trabue J	.32	.28	.27	.25		.42	.00	.15	.45	.43	.31	.29	.63	.40	.33	.29	.25	.36
Trabue K	.50	.43	.27	.24	.42		.14	.10	.50	.27	.23	.31	.65	.24	.07	.09	—.02	.30
Woody Mult.	.24	.23	.04	.44	.00	.14		.18	.38	—.08	.21	.18	.36	.35	.18	.39	.29	.43
Woody Div.	—.02	.13	.27	.12	.15	.10	.18		.36	.19	.23	.17	.30	.35	.37	.34	.08	.38
Opposites	.50	.51	.41	.43	.45	.50	.38	.36		.52	.50	.46	.70	.50	.28	.47	.26	.44
Easy Direc.	.33	.26	.10	.12	.43	.27	—.08	.19	.52		.36	.25	.52	.27	.15	.18	.04	.34
Mixed Rel.	.32	.29	.05	.10	.31	.23	.21	.23	.50	.36		.26	.54	.18	.24	.18	.26	.36
Average	.34	.33	.22	.24	.29	.31	.18	.17	.46	.25	.26		.55	.39	.25	.32	.17	
Composite	.69	.67	.50	.54	.63	.65	.36	.30	.70	.52	.54	.55		.68	.38	.55	.23	
Teachers' Rank	.43	.47	.49	.60	.40	.24	.35	.35	.50	.27	.18	.39	.68		.50	.90	.30	
Marks First 6 Years	.24	.20	.29	.36	.33	.07	.18	.37	.28	.15	.24	.25	.38	.50		.49	.57	
Marks First Year Jr. H. S.	.32	.37	.38	.56	.29	.09	.39	.34	.47	.18	.18	.32	.55	.90	.49		.34	
Youth	.27	.15	—.03	.27	.25	—.02	.29	.08	.26	.04	.26	.17	.23	.30	.57	.34		
With Itself 1916	.56	.52	.32	.52	.36	.30	.43	.38	.44	.34	.36							

TABLE P

Average Correlation of Each Test With Every Other Test 1916-1917

	Vis. Vocab.	Reading	Composition	Spelling	Trabue B-J	Trabue C-K	Woody Mult.	Woody Div.	Opposites	Easy Direc.	Mixed Rel.	Average	Composite	Teachers' Rank	Marks First 6 Yrs.	Marks First Year Jr. H. S.	Youth	With Itself After 1 Year
Vis. Vocab.		.55	.33	.30	.27	.46	.08	.01	.42	.39	.39	.32	.71	.43	.19	.33	.15	.56
Reading	.55		.28	.23	.30	.39	.11	.07	.30	.36	.30	.29	.65	.47	.21	.40	.16	.52
Composition	.33	.28		.16	.18	.31	.10	.15	.32	.15	.06	.21	.50	.43	.25	.35	.06	.32
Spelling	.30	.23	.16		.22	.19	.35	.19	.30	.20	.10	.22	.53	.48	.38	.44	.27	.52
Trabue B-J	.27	.30	.18	.22		.35	.06	.05	.35	.25	.20	.22	.54	.29	.24	.22	.18	.36
Trabue C-K	.46	.39	.31	.19	.35		.09	.02	.41	.25	.34	.28	.62	.29	.10	.20	.02	.30
Woody Mult.	.08	.11	.10	.35	.06	.09		.29	.20	-.16	.09	.12	.31	.30	.23	.33	.28	.43
Woody Div.	.01	.07	.15	.19	.05	.02	.29		.13	.13	.09	.11	.28	.27	.28	.30	.13	.38
Opposites	.42	.30	.32	.30	.35	.41	.20	.13		.41	.36	.32	.59	.43	.22	.36	.11	.44
Easy Direc.	.39	.36	.15	.20	.25	.25	-.16	.13	.41		.33	.23	.55	.31	.12	.23	.04	.34
Mixed Rel.	.39	.30	.06	.10	.20	.34	.09	.09	.36	.33		.23	.54	.20	.11	.17	.17	.36
Average	.32	.29	.21	.22	.28	.28	.12	.11	.32	.23	.23		.53	.34	.21	.30	.15	
Composite	.71	.65	.50	.53	.54	.62	.31	.28	.59	.55	.54	.53		.67	.33	.56	.22	
Teachers' Rank	.43	.47	.43	.48	.29	.29	.30	.27	.43	.31	.20	.34	.67		.50	.90	.30	
Marks First 6 Years	.19	.21	.25	.38	.24	.10	.23	.28	.21	.12	.11	.21	.33	.50		.49	.57	
Marks First Year Jr. H. S.	.33	.40	.35	.44	.22	.20	.33	.30	.36	.23	.17	.30	.56	.90	.49		.34	
Youth	.15	.16	.06	.27	.18	.02	.28	.13	.11	.04	.17	.15	.22	.30	.57	.34		
With Itself After 1 Year	.56	.52	.32	.52	.36	.30	.43	.38	.44	.34	.36							

9. Corrected Coefficients of Correlation

Since chance inaccuracies in the paired measures correlated do not render each other harmless but tend to produce zero correlation, it is necessary to correct the raw coefficients. As either the same tests or tests similar to those given in 1916 had been repeated one year later, the two independent measures necessary for this correction for "attenuation" due to chance errors, are at hand. By utilizing the raw Pearson coefficients of correlation of Table Q, it is possible to present the corrected coefficients in Table R. The formula[1] used is

$$r_{pq} = \frac{\sqrt{(r_{p1q2})\ (r_{p2q1})}}{\sqrt{(r_{p1p2})\ (r_{q1q2})}}$$

If Visual Vocabulary and Reading are the measures to be related, let A equal the former and B the latter. Let p be a series of exact measures of A, and q be the related series of exact measures of B. Let r_{pq} be the coefficient of correlation of A and B, obtainable from the two series p and q. r_{pq} is thus, according to this theory that errors are due to chance errors in the data, the required true coefficient. Let p_1 and p_2 be two independent series of measures of A, and q_1 and q_2 two independent series of measures of B. Let r_{p1q2} be the correlation when the first measure of A and the second measure of B are used, and r_{p2q1} be the correlation when the second measure of A and the first measure of B are used. Let p_1p_2 be the correlation between the two measures of A, and q_1q_2 the correlation between the two measures of B. Of course a test could be split and the odd responses, for example, be correlated against the even, but this was not necessary here as the eleven tests were repeated after one year.

In Table R, since some raw coefficients were either zero or negative, there are some coefficients wanting. Also, since in some cases, the p_1p_2 and q_1q_2 were very low, some corrected coefficients are 1+. Due to this and to the additional fact that the practical administrator must depend on raw coefficients, more use has been made in this study of the raw than of the corrected coefficients.

[1] Thorndike, E. L., *Mental and Social Measurements*, p. 179, 1913 edition.

TABLE Q

RAW COEFFICIENTS OF CORRELATION

		Vis. Vocab. '16	Vis. Vocab. '17	Reading '16	Reading '17	Compos. '16	Compos. '17	Spelling '16	Spelling '17	Trabue B '16	Trabue C '16	Trabue J '17	Trabue K '17
Vis. Vocab....	'16		.56		.32		.28		.37			.31	.49
Vis. Vocab....	'17	.56				.43		.28		.33	.36		
Reading	'16		.56		.52	.52	.26		.19			.26	.32
Reading	'17	.32				.41		.13		.19	.30		
Composition ..	'16		.43		.41		.32		.26			.38	.51
Composition ..	'17	.28		.26				.23		.23	.27		
Spelling	'16		.28		.13		.23		.52			.35	.21
Spelling	'17	.37		.19		.26				.16	.12		
Trabue B	'16		.33		.19		.23		.16			.36	.22
Trabue C	'16		.36		.30		.27		.12			.29	.30
Trabue J	'17	.31		.26		.38		.35		.36	.29		
Trabue K	'17	.49		.32		.51		.21		.22	.30		
Woody Mult..	'16		.15		.06		.06		.21			.03	—.01
Woody Mult..	'17	.08		.16		.20		.25		.21	.20		
Woody Div...	'16		—.12		—.04		.09		.22			.04	.18
Woody Div...	'17	.12		.15		.18		.31		.01	.03		
Opposites	'16		.25		.29		.32		.22			.35	.18
Opposites	'17	.58		.48		.48		.45		.44	.43		
Easy Direc....	'16		.35		.51		.27		.31			.31	.38
Easy Direc....	'17	.34		.17		.24		.19		.09	.22		
Mixed Rel. ..	'16		.37		.31		.12		.14			.23	.21
Mixed Rel. ..	'17	.15		.28		.12		.12		.17	.22		

TABLE Q—*Continued*

RAW COEFFICIENTS OF CORRELATION

		Woody Mult. '16	Woody Mult. '17	Woody Div. '16	Woody Div. '17	Opposites '16	Opposites '17	Easy Direc. '16	Easy Direc. '17	Mixed Rel. '16	Mixed Rel. '17
Vis. Vocab......	'16		.08		.12		.58		.34		.15
Vis. Vocab......	'17	.15		—.12		.25		.35		.37	
Reading	'16		.16		.15		.48		.17		.28
Reading	'17	.06		—.04		.29		.51		.31	
Composition	'16		.20		.18		.48		.24		.12
Composition	'17	.06		.09		.32		.27		.12	
Spelling	'16		.25		.31		.45		.19		.12
Spelling	'17	.21		.22		.22		.31		.14	
Trabue B	'16		.21		.01		.44		.09		.17
Trabue C	'16		.20		.03		.43		.22		.22
Trabue J	'17	.03		.04		.35		.31		.23	
Trabue K	'17	.01		.18		.18		.38		.21	
Woody Mult....	'16		.43		.37		.20		—.06		.14
Woody Mult....	'17	.31		.15		.08		.16			
Woody Div.....	'16		.31		.38		.30		.22		.11
Woody Div.....	'17	.37		.13		.10		.10			
Opposites	'16		.15		.13		.44		.32		.24
Opposites	'17	.20		.30		.59		.44			
Easy Direc......	'16		.08		.10		.34		.27		
Easy Direc......	'17	—.06		.22		.32				.23	
Mixed Rel.	'16		.16		.11		.10		.24		.44	.23	.36
Mixed Rel.	'17	.14		.24		.27					

TABLE R

CORRECTED COEFFICIENTS OF CORRELATION

	Vis. Vocab.	Reading	Composition	Spelling	Trabue B-J	Trabue C-K	Woody Mult.	Woody Div.	Opposites	Easy Direc.	Mixed Rel.
Vis. Vocab......		.79	.83	.60	.71	1.02*	.22		.76	.79	.52
Reading	.79		.80	.30	.51	.79	.20		.76	.70	.68
Composition	.83	.80		.59	.86	1.19*	.30	.37	1.04*	.77	.35
Spelling	.60	.30	.59		.54	.41	.49	.73	.65	.59	.30
Trabue B-J......	.71	.51	.86	.54		.75	.21	.05	.99	.48	.61
Trabue C-K.....	1.02*	.79	1.19*	.41	.75		.39	.22	.78	.90	.68
Woody Mult.....	.22	.20	.30	.49	.21	.39		.84	.39		.38
Woody Div......			.37	.73	.05	.22	.84		.49	.42	.27
Opposites	.76	.76	1.04*	.65	.99	.78	.39	.49		1.13*	.86
Easy Direc......	.79	.70	.77	.59	.48	.90		.42	1.13*		.71
Mixed Rel.	.52	.68	.35	.30	.61	.68	.38	.27	.86	.71	

10. SELECTION OF TESTS

The present evaluation of tests involves two chief questions: First, the evaluation of individual tests for the purpose of educational prognosis, and, second, the combination of tests to use in such an experiment as this study has recorded. Of the seven standards proposed, pages 21 and 22, standards 2 and 4 are considered of most worth, and the ranking of the tests as given under two and four in Table U is believed to be more nearly correct than that of any of the other combinations of standards. In every combination of standards presented in Tables U and V, Reading, Visual Vocabulary, Opposites, and Spelling come in the first division of the whole group of tests. The practical administrator can add the Completion and the Arithmetic tests to this list of four tests if he desires to extend his testing beyond seventy-five minutes.

If all of these tests correlated +1. with each other, there would be no need of giving more than one of them. Evidently such a correlation would indicate that the tests measured the same traits. Since nearly all of these are language tests, it is to be expected that the Arithmetic tests would have a low correlation with the composite and a low correlation with every other test. This fact that the Arithmetic tests, which have been found to

TABLE S

CORRELATION OF EACH TEST WITH STANDARDS PROPOSED
FOR EVALUATING TESTS

	I Correlation Between Repeated Tests	II Correlation With Its Each Test Composite		III Average Correlation of Each Test With Every Other Test		IV Correlation Each Test With Teachers' Rankings		V Correlation Each Test With All Marks During 1st Year Junior H. S.		VI Correlation Each Test With All Marks Before Junior H. S.		VII Correlation Each Test With Youth Within School Grade		VIII Economy of Time in Giving Test Ranks	IX Estimated Economy of Time in Scoring Test Ranks
	1916-17	1916	1917	1916	1917	1916	1917	1916	1917	1916	1917	1916	1917	Ranks	Ranks
Visual Vocabulary ...	.56	.73	.69	.31	.34	.44	.33	.34	.32	.15	.24	.04	.27	10	9
Reading	.52	.63	.67	.26	.33	.47	.47	.43	.37	.22	.20	.17	.15	10	10
Composition	.32	.51	.50	.20	.22	.34	.49	.38	.38	.21	.29	.15	—.03	10	11
Spelling	.52	.53	.54	.21	.24	.37	.60	.32	.56	.40	.36	.27	.27	8	8
Trabue B-J	.36	B.45	J.63	.16	.29	B.18	J.40	B.16	J.29	B.16	J.33	.12	.25	4.5	6.5
Trabue C-K	.40	C.59	K.65	.26	.31	C.38	K.24	C.31	K.09	C.13	K.07	.07	—.02	4.5	6.5
Woody Mult.	.43	.26	.36	.07	.18	.25	.35	.28	.39	.29	.18	.27	.29	6.5	4.5
Woody Div.	.38	.26	.30	.06	.17	.20	.35	.27	.38	.19	.37	.19	.08	6.5	4.5
Opposites	.44	.49	.70	.19	.46	.36	.50	.26	.47	.16	.28	—.03	.26	1	1
Easy Direc.	.34	.58	.52	.22	.25	.35	.27	.29	.18	.10	.15	.04	.04	2	2
Mixed Rel.	.36	.55	.54	.20	.26	.22	.18	.17	.18	—.02	.24	.09	.26	3	3

have a positive correlation with desirable traits, do have a low average correlation with every other test, probably indicates that they measure some abilities that the others do not measure. If such is the case, the test that has the lowest correlation with every other test should be ranked one, and the test with the highest average correlation, ranked eleven. Such a ranking has been made in standard 3 of Table S as ranked in Table V.

TABLE T

RANKING OF TESTS BY ALL STANDARDS

Rank by Standards:	I	II	III	IV	V	VI	VII	VIII	IX
Visual Vocab. ...	1	1	1	3	5.5	8	7	10	9
Reading	2.5	2	2	1.5	2	6.5	3	10	10
Composition	10	9	9	4	3	3	9	10	11
Spelling	2.5	7.5	7	1.5	1	1	2	8	8
Trabue B-J	7.5	7.5	8	9	9.5	4	5	4.5	6.5
Trabue C-K	11	3	4	6	9.5	11	10.5	4.5	6.5
Woody Mult.	5	10	10	8	5.5	5	1	6.5	4.5
Woody Div.	6	11	11	10	7	2	6	6.5	4.5
Opposites	4	4	3	5	4	6.5	8	1	1
Easy Direc.	9	5.5	5	7	8	10	10.5	2	3
Mixed Rel.	7.5	5.5	6	11	11	9	4	3	2

TABLE U

Rank by Standards:	II and IV	I to IV	I to V	I to VII
Visual Vocab.	2	1	2	3
Reading	1	2	1	2
Composition	6	8.5	7	6
Spelling	4	4	3	1
Trabue B-J	8.5	8.5	10	7
Trabue C-K	4	5	6	10.5
Woody Mult.	10	9	8	5
Woody Div.	11	10	11	9
Opposites	4	3	4	4
Easy Direc.	7	6	5	10.5
Mixed Rel.	8.5	7	9	8

TABLE V

Rank by Standards:	I to IV	I to V	I to VII
Visual Vocab.	1.5	3	3
Reading	1.5	2	2
Composition	6	5	5
Spelling	3	1	1
Trabue B-J	8.5	9.5	8
Trabue C-K	8.5	9.5	11
Woody Mult.	5	6	4
Woody Div.	7	7	7
Opposites	4	4	6
Easy Direc.	9	8	10
Mixed Rel.	10	11	9

The results, whether the grouping of standards be one to four, one to five, or one to seven, call attention to the fact that in a combination of tests for educational prognosis, the Arithmetic tests hold a relatively higher place than they do when considered as in Table **T**.

In selecting a test, two other standards must be taken into account with those already considered: Economy of the pupil's time in taking the test, and economy of the administrator's time in scoring it. Standard 8 in Table S indicates the relative time consumed by the pupils in taking the tests, and standard 9 indicates, likewise, the relative time necessary to score the tests. In any practical experiment, these two standards must be considered. For example, regardless of the importance of Composition as a test, it is very difficult to use it. The variability in grading even by skilled persons using an objective scale is so great that the same paper must be read by three or more persons and their scores averaged, in order to secure an approximately accurate grade. Next to Composition in time required both for taking the test and for scoring it, come Visual Vocabulary and Reading. However, in each of these cases the scoring requires the reading of only one person, and this score can be approximately accurate. In speed of giving and ease of scoring, Opposites, Mixed Relations, and Easy Directions are easily at the head of the list.

11. Correlation of Combinations of Tests with Teachers' Ranking and with Composite of Eleven Tests

For the administrator who, for any reason, does not wish to use all eleven tests considered in this study, it has been pointed out that Visual Vocabulary, Reading, Opposites, Spelling, Completion Tests, Woody Multiplication are the tests he can use to greatest advantage. Table W indicates the success that would have been met with in this study if these tests had been used in the order mentioned. In reading this table it should be held in mind that the correlation of the composite of all eleven tests with the Teachers' Ranking was, for the 1916 tests, .66, and for the 1917 tests, .68. The correlation of Visual Vocabulary with the composite of eleven tests in 1916 was .73, in 1917, .69; with the

composite of the Teachers' Ranking 1916, .44, 1917, .43. The
corresponding figures for Reading are .63, .67, .47, .47. The
average correlation of Visual Vocabulary with the composite
then is .71, and with Teachers' Ranking, .43. Reading is some-
what lower, with averages of .65 and .47. Therefore, the aver-
age correlations of Visual Vocabulary and Reading are, .68 with
the composite and .45 with the Teachers' Ranking. However,
when Visual Vocabulary and Reading are combined, as in Table
W, the correlation is not that of the average of correlations, .68
with the composite and .45 with Teachers' Ranking, but is raised
in 1916 to .77 and .54, i.e., the combination has raised the corre-
lation about .10 in each case. The Completion Tests, which
probably measure somewhat the same qualities as Reading and
Visual Vocabulary, could have been used so far as their corre-
lations with the composite are concerned. Thus J and K com-
bined have a correlation of .76 with the 1917 composite. This
is as high as that of Reading and Visual Vocabulary combined,
and these tests can be given quicker and scored more easily than
can Reading and Visual Vocabulary. However, instead of hav-
ing a correlation of .54 with Teachers' Ranking, as Reading and
Visual Vocabulary have in 1916, the Completion Tests when
combined have a correlation of .36 with Teachers' Ranking.
Since teacher-judgments must play so large a part in a practi-
cal experiment, the reason for using Visual Vocabulary and
Reading instead of the Completion Tests is apparent. The aver-
age correlations of Reading, Visual Vocabulary, and Opposites
with the composites of 1916 and of 1917 with the Teachers'
Rankings are .65 and .44. But when these three tests are com-
bined as one test the correlations are raised from .65 and .44 to
.82 and .57 in 1916. If to the three tests just mentioned Spell-
ing is added, the average correlation for all four tests for both
years is .62 with the composite and .46 with Teachers' Ranking;
while the four tests combined as one test have correlations of
.88 and .64 for 1916 and 1917 combined. These four tests then
lack only .03 of having as high a correlation with Teachers'
Ranking as do the whole eleven tests, and, at the same time,
they have a correlation with the composite of .87. On the basis
of this experiment, this is the result that may be expected from

TABLE W

CORRELATION OF COMBINATIONS OF TESTS WITH TEACHERS' RANKING AND COMPOSITE OF ELEVEN TESTS

	Combination	With Teachers' Ranks		With Composite of Eleven Tests	
		1916	1917	1916	1917
	Reading and Vis. Vocab.	.54	.50	.77	.76
	" " " " Opposites "	.57	.54	.82	.82
	" " " " " " Spelling "	.62	.66	.88	.87
	" " " " " " " " Trabue B & C	.61	.63	.91	.93
	" " " " " " " " " " J & K	.66	.65	.93	.92
	" " " " " " " " " " Woody Mult. " "				

TABLE X

CORRELATION OF COMBINATIONS OF TESTS WITH TEACHERS' RANKING AND COMPOSITE OF ELEVEN TESTS

	Combination	With Teachers' Ranks		With Composite of Eleven Tests	
		1916	1917	1916	1917
	Reading and Vis. Vocab.	.54	.50	.77	.76
	" " " " Opposites "	.57	.54	.83	.82
	" " " " " " Trabue B & C	.57	.53	.85	.89
	" " " " " " " " J & K	.58	.52	.87	.90
	" " " " " " " " " " Easy Direc.	.62	.61	.93	.94
	" " " " " " " " " " " " Spelling				

a little less than one and one quarter hours' testing. A further refinement, as is shown in Tables W and X, can be had by using the additional tests indicated.

According to the classification by these four tests, Reading, Visual Vocabulary, Opposites, and Spelling, had the classes in 1916 been composed of thirty pupils, there would have been, according to the Teachers' Rankings one year later, only eight displacements. That is, when all temporary illnesses on the part of the pupils, ranging from "bad colds" through contagious diseases to a month in the hospital, all fortunes or misfortunes in the home life, barring the withdrawal of the pupil from school, all the changing physical conditions and varying interests in boys of eleven to thirteen—when all these and a score of others that might be enumerated are considered, the use of these tests in one and one quarter hours' testing at the beginning of the year would have agreed with the classification of the teachers after teaching the pupils one year in ninety per cent of all cases.

12. Conclusion

1.   In this study, academic success in the first year of junior high school was more successfully predicted by a group of standardized tests than by all previous school marks or age or teachers' estimates.

2.   The tests in the order of their importance for the purposes of this study, when the administration and scoring of the tests are considered, have been found to be: Reading, Visual Vocabulary, Opposites, Spelling, Completion Tests, Arithmetic Tests, Easy Directions, Mixed Relations, and Composition.

TABLE Y

Score by Eleven Tests—February, 1916

Pupil	Spelling	Comp.	Woody ×	Woody ÷	Trabue B	Trabue C	Reading A2	Opposites	Mixed Rel.	Easy Direc.	Vis. Voc.
Total Possible Score	100	93	39	36	20	20	48	20	20	20	43
1	100	30.2	33	36	12	10	13	13	2	9.5	27
2	96	23.4	32	34	17	10.5	16.5	18	18	17	42
3	98	22.2	32	23	14	12	11	17	10	11	26
4	98	48.6	36	36	16	16	15	18	17	16	33
5	96	27.5	34	24	11	11	14.5	14	5	18	23
6	100	24.5	33	24	14	13	12.5	18	18	11	26
7	98	37.8	30	33	13	14	14	14	18	17	28
8	96	37.7	32	32	11	6	14.5	18	16	19	32
9	92	44.7	32	29	13	14	14.5	18	19	20	33
10	100	40.6	36	33	11	11	17	14	18	16	32
11	98	43.2	37	32	13	16	16.5	10	16	12	35
12	94	36	33	35	11	14	14	13	6	20	18
13	88	29.2	29	27	4	11	2.5	16	9	14	24
14	100	27.4	32	33	11	15	11.5	19	14	20	28
15	96	34.5	30	28	13	11	16	19	7	15	32
16	100	33.6	35	33	11	14	15	18	19	20	30
17	92	34.8	29	26	14	15	14.5	16	17	16	39
18	84	33.2	22	24	14	10	13	16	4	12	19
19	92	27.4	37	35	14	12	11	13	15	11	26
20	100	39.2	35	29	17	18	15	19	20	16	40
21	58	25.2	33	27	8	10	5.5	14	11	11	17
22	98	31	28	29	11	13	13.5	16	19	18	35
23	100	34	36	33	15	13	13	18	9	13	43
24	96	26.5	27	33	12	15	12	20	19	19	34
25	98	50	32	33	11	14	15	20	19	14	38
26	74	37.5	33	30	16	13	12	20	13	14	30
27	96	39.6	35	35	13	10	10.5	18	4	14	26
28	96	32	32	18	15	16	15	16	17	13	28
29	98	43.5	29	27	12	17	15	19	2	14	26
30	98	35.2	35	34	13	11	8	17	13	11	22
31	98	33	34	36	18	12	9.5	18	14	13	21
32	98	32.2	32	33	14	14	16	20	5	16	30
33	98	21.2	31	31	14	15	18	18	20	18	38
34	80	30	30	33	13	12	10	17	15	13	29
35	100	33.5	31	28	15	15	10	15	13	11	30
36	96	28	23	25	11	10	10.5	19	17	16	32
37	96	44.5	36	33	15	17	15.5	20	9	19	37
38	94	25.2	36	34	13	12	15	13	13	13	20
39	94	28.5	31	25	17	10	15.5	17	9	17	27
40	86	30.8	25	29	12	13	14	19	7	16	30
41	100	30.6	31	31	14	8	17.5	16	11	18	32
42	98	40.6	34	31	15	11	12	18	19	17	28
43	86	46	33	27	10	13	11	18	17	11	26
44	100	44	28	31	14	19	20	20	18	20	34
45	98	26.4	37	33	8	10	7.5	17	13	13	28
46	98	40.8	37	35	13	15	17.5	20	18	15	38
47	92	46	33	32	12	16	16	16	18	14	39
48	96	38.5	37	34	10	12	10	19	9	7	33
49	100	42.5	33	28	12	13	12.5	17	16	14	34
50	86	17.5	11	32	12	12	14.5	15	15	13	34
51	98	56.6	31	30	15	12	12	17	5	16	38
52	92	47.8	33	27	14	15	15	16	11	11	31

TABLE Y—*Continued*

SCORE BY ELEVEN TESTS—FEBRUARY, 1916

Pupil	Spelling	Comp.	Woody ×	Woody ÷	Trabue B	Trabue C	Reading A2	Opposites	Mixed Rel.	Easy Direc.	Vis. Voc.
Total Possible Score	100	93	39	36	20	20	48	20	20	20	43
53............	96	34	36	27	12	10	12.5	16	7	11	25
54............	100	40.8	31	32	11	10	9	19	8	18	36
55............	64	34.8	27	26	12	15	11.5	18	17	12	27
56............	94	36.6	29	23	14	14	14	20	16	13	28
57............	94	38.8	33	27	14	13	16.5	18	19	14	39
58............	98	34.2	30	32	8	12	16	12	17	19	32
59............	100	33.4	34	35	16	12	17	16	7	13	28
60............	96	28.5	33	33	10	14	14	10	12	13	25
61............	96	22.6	37	28	16	16	14	19	16	11	32
62............	100	32.6	36	28	13	12	12	19	3	14	23
63............	98	54	36	30	16	16	19.5	20	20	20	42
64............	64	23.7	32	29	13	10	11.5	17	12	10	21
65............	94	36.6	36	30	10	12	8	17	11	8	22
66............	96	35.8	33	31	8	12	16	19	15	15	31
67............	96	26.5	33	35	12	12	9.5	11	11	9	19
68............	96	26.7	28	31	12	12	15.5	15	17	14	31
69............	100	20.2	33	31	13	14	12.5	19	17	20	33
70............	92	34.5	29	35	11	12	10.5	11	5	16	34
71............	92	26	34	33	12	12	11	20	14	14	20
72............	98	30.4	34	34	11	9	13	15	5	9	28
73............	98	31.7	33	33	17	11	15.5	17	11	16	24
74............	98	46.5	33	31	14	13	14	18	6	16	36

TABLE Z

SCORE BY ELEVEN TESTS—FEBRUARY, 1917

Pupil	Spelling	Comp.	Woody ×	Woody ÷	Trabue J	Trabue K	Reading A2	Opposites	Mixed Rel.	Easy Direc.	Vis. Voc.
Total Possible Score	100	93	20	15	14	14	48	20	20	20	170
1............	86	52	16	13	5	8	22	16	10	14	97
2............	92	44.7	15	13	6	8	29	20	2	17	126
3............	68	51.2	14	13	6	5	25	16	12	19	119
4............	90	60.5	19	14	9	12	37	20	17	18	122
5............	84	39.2	17	12	5	2	38	19	8	15	131
6............	78	43.5	13	14	8	5	32	20	17	19	114
7............	78	58.2	15	12	8	6	33	20	9	20	108
8............	96	43.7	17	14	7	10	40	20	18	19	133
9............	98	53.5	17	11	7	8	38	20	17	16	142
10............	74	45.2	16	14	5	3	35	20	15	16	125
11............	80	55.5	14	14	6	7	28	20	13	16	133
12............	82	41.5	14	14	8	6	29	20	19	20	100
13............	76	41	11	12	5	5	16	19	3	17	51
14............	90	44.5	16	13	4	6	31	20	10	20	109
15............	80	47.5	16	13	6	6	26	20	13	18	108

TABLE Z—*Continued*

16............	94	53.7	16	13	8	9	37	20	17	20	117
17............	76	46	14	9	8	8	39	20	17	20	126
18............	54	41.5	12	11	7	5	15	17	6	11	100
19............	88	38.5	15	13	8	5	25	18	13	20	123
20............	94	42	18	13	10	9	37	20	18	20	162
21............	40	46	16	13	4	6	30	16	15	12	35
22............	66	47.5	14	13	8	10	33	20	12	18	140
23............	96	49.7	16	14	8	8	29	20	5	20	130
24............	80	50.5	14	13	7	5	30	19	10	20	116
25............	94	45.2	18	14	7	8	39	20	20	20	144
26............	70	45	13	11	8	12	36	20	17	20	163
27............	72	49	16	14	8	10	27	20	18	19	106
28............	78	47	16	7	6	8	27	20	19	19	146
29............	88	48	15	13	10	10	37	20	5	19	120
30............	88	56.5	17	13	9	4	26	20	18	18	100
31............	100	48.2	19	13	5	6	23	20	15	17	106
32............	84	58.7	17	14	7	4	39	20	18	20	121
33............	88	47.5	11	12	8	7	31	20	19	18	129
34............	46	40.5	15	13	8	8	30	20	17	16	97
35............	90	38.2	16	11	8	6	20	20	7	13	106
36............	96	39.5	14	13	4	5	31	20	17	19	96
37............	88	58.7	17	14	7	8	42	20	12	17	139
38............	84	47.2	17	12	4	6	35	19	12	11	117
39............	74	54.5	12	13	9	6	36	20	16	18	117
40............	54	43.2	14	11	4	7	30	20	16	20	118
41............	92	43.2	15	13	7	8	37	20	16	20	145
42............	88	45.7	15	14	9	7	34	20	19	20	125
43............	74	44	16	12	6	7	38	20	11	19	136
44............	80	60.5	14	14	10	10	36	20	19	19	143
45............	80	47.2	14	13	6	7	31	20	18	20	123
46............	88	47.2	16	12	6	8	35	20	17	20	142
47............	82	53	17	13	7	12	39	20	19	19	144
48............	90	44.2	17	12	8	4	20	20	15	20	117
49............	96	48.7	15	14	6	6	29	20	16	17	136
50............	92	57.7	15	13	6	6	36	20	10	19	111
51............	92	56	14	10	6	8	38	20	16	18	158
52............	90	42	16	13	5	5	32	20	8	18	121
53............	70	39	16	13	6	7	29	16	5	14	118
54............	94	45.7	15	13	10	9	28	19	12	20	129
55............	32	42	15	12	6	5	29	17	12	17	115
56............	88	48.7	14	11	11	6	32	19	17	20	123
57............	76	50.5	15	13	8	6	33	20	17	20	144
58............	96	42	18	13	8	7	20	20	17	19	113
59............	92	37.7	16	13	9	8	34	20	18	20	142
60............	92	39.7	16	13	6	7	35	11	19	17	129
61............	86	42.7	17	13	6	5	17	19	19	12	104
62............	90	49.2	15	13	7	5	34	20	16	18	133
63............	92	52.7	15	14	11	10	41	20	20	20	154
64............	72	48	13	14	5	4	25	19	16	18	103
65............	94	47	14	13	6	6	22	17	7	7	109
66............	88	55.5	14	14	9	5	34	19	13	19	120
67............	66	41.7	14	14	5	3	25	18	16	20	112
68............	78	40	15	12	7	6	19	20	18	20	137
69............	98	48.2	18	12	8	5	29	20	17	16	140
70............	70	44.2	12	14	5	6	23	20	17	20	73
71............	66	43.2	17	13	6	4	31	20	18	17	87
72............	80	40	9	12	7	5	29	18	12	15	106
73............	94	45	17	12	7	5	37	20	19	19	138
74............	98	51.2	14	12	10	9	32	20	14	20	118

TABLE AA

RANKING BY TEACHERS AFTER TEACHING PUPILS

Individual	Ranking by age, youngest pupil ranked 1, oldest ranked 74	Ranking by school marks, 1st 6 years	Ranking by school marks, 1st yr. Jr. H.S.	Teacher No. 1	Teacher No. 2	Teacher No. 3	Teacher No. 4
1	13.5	4.5	47	42	43	56	66
2	33	70	46	53	50	57	9
3	45.5	65	61	70	51	45	45
4	13.5	15.5	29	4	27	23	63
5	8.5	55.5	50	72	73	22	46
6	2.5	22.5	56	40	62	70	37
7	35.5	55.5	62	52	64	53	53
8	28.5	15.5	33	26	35	11	31
9	17	34	11	20	15	35	12
10	45.5	55.5	25	29	21	54	18
11	31	34	20.5	18	31	3	19
12	2.5	9	26.5	33	30	13	44
13	62	70	64	68	67	51	43
14	53.5	34	14	11	29	44	6
15	62	34	66	51	63	40	65
16	69	34	29	3	16	26	2
17	70	70	68	41	66	19	36
18	72	65	71	62	56	63	74
19	40	55.5	37.5	45	54	62	39
20	7	4.5	5	8	6	12	5
21	74	65	52.5	49	57	69	35
22	52	15.5	39	44	65	18	26
23	71	22.5	8	7	8	16	1
24	40	55.5	48	48	28	31	61
25	10.5	15.5	3.5	1	2	4	8
26	49.5	55.5	43	37	32	27	33
27	57	55.5	60	65	58	58	55
28	69	55.5	63	67	52	64	60
29	59	43	18	15	23	17	20
30	28.5	4.5	12.5	13	20	28	57
31	45.5	55.5	24	27	33	5	28
32	13.5	1	8	12	9	30	11
33	10.5	9	22.5	34	37	74	25
34	40	70	72	66	72	47	67
35	64.5	55.5	74	69	74	71	51
36	59	47	45	55	45	68	40
37	56	4.5	16	25	3	6	10
38	13.5	9	35	47	42	46	16
39	33	34	15	36	22	7	41
40	69	70	73	74	60	66	50
41	49.5	34	42	17	18	72	73
42	20.5	22.5	70	38	46	38	68
43	20.5	55.5	55	64	68	52	62
44	40	55.5	31.5	30	7	2	30
45	49.5	70	65	54	61	34	38
46	59	74	40.5	24	36	41	72
47	5	43	28	35	26	36	24
48	5	15.5	8	5	10	14	7
49	45.5	43	20.5	22	34	33	54
50	55	55.5	8	9	14	9	15
51	17	15.5	36	32	12	20	56
52	40	43	26.5	31	19	24	29
53	35.5	43	67	63	55	61	48
54	24.5	34	52.5	43	39	37	70
55	40	55.5	34	56	59	25	34
56	40	34	54	39	40	21	52
57	24.5	22.5	22.5	28	38	32	14
58	8.5	55.5	1	2	4	49	4
59	2.5	27	12.5	14	25	8	17
60	53.5	43	57	57	70	43	71
61	20.5	34	44	59	41	42	23

TABLE AA—*Continued*

RANKING BY TEACHERS AFTER TEACHING PUPILS

Individual	Ranking by age, youngest pupil ranked 1, oldest ranked 74	Ranking by school marks, 1st 6 years	Ranking by school marks, 1st year jr.h.s	Teacher No. 1	Teacher No. 2	Teacher No. 3	Teacher No. 4
62	1	2	31.5	21	24	15	32
63	49.5	9	3.5	23	1	1	27
64	73	22.5	58.5	58	44	73	42
65	24.5	27	37.5	46	53	55	22
66	62	22.5	8	6	5	10	3
67	24.5	9	49	61	48	29	69
68	20.5	15.5	69	71	69	60	64
69	17	27	30	16	17	50	21
70	66	55.5	51	60	71	67	58
71	64.5	70	40.5	50	47	39	47
72	28.5	34	58.5	73	49	59	59
73	28.5	43	2	10	11	48	13
74	33	15.5	19	19	13	65	49